LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 74
ottobre 1974
anno VII
volume XIII

L'olografia nel campo del restauro

La diagnosi tempestiva e minuziosa di difetti nei dipinti e nelle sculture attuata con i metodi interferometrici dà un valido ausilio alla conservazione delle opere d'arte

di Franco Gori e Giovanni Urbani

Dipinti, statue e monumenti, come del resto ogni altra componente « nobile » dell'ambiente, risultano sempre più esposti a vistosi e velocissimi fenomeni di deterioramento. Naturalmente l'inquinamento è una delle cause di questa situazione, ma non la sola e forse nemmeno la più importante. Per paradossale che possa sembrare, il nostro patrimonio storico-artistico va in pezzi soprattutto perché dal momento stesso in cui si è cominciato a considerarlo come un patrimonio (cioè come qualcosa di irremissibilmente concluso, formato una volta per tutte e non più ripetibile, e appunto per questo prezioso) è stato praticamente abbandonato a se stesso, nel timore (del resto sacrosanto) che porvi mano con riparazioni o rifacimenti significasse privarlo del suo valore principale: l'autenticità. Non per nulla il Romanticismo, cioè il momento culturale in cui si forma la nostra concezione della storia, aveva identificato nel rudere e nel frammento l'espressione più pura del fenomeno estetico.

È così accaduto che, sebbene i tentativi di porre su basi scientifiche il restauro delle opere d'arte datino dall'inizio di questo secolo, essi abbiano finora ricevuto un sufficiente impulso solo per quanto riguarda la definizione dei criteri (e la messa a punto dei relativi mezzi tecnici) con cui effettuare le cosiddette operazioni di recupero o di restituzione dei caratteri originali delle opere; mentre non si possono contare che scarse innovazioni per quanto riguarda le tecniche di consolidamento e di stabilizzazione dei materiali. È dunque evidente che per la loro dipendenza dal criterio del rispetto dell'autenticità, alla tecnica del restauro e alle sussidiarie indagini di laboratorio è stato assegnato un raggio d'azione che oggi risulta troppo ristretto rispetto ai problemi posti dal progressivo deterioramento dell'ambiente e dalla sempre minore resistenza offerta dai materiali e dalle strutture delle opere d'arte ai naturali processi di invecchiamento.

Immagine ripresa attraverso un ologramma ordinario di un dipinto su tavola (la « Santa Caterina » di Pier Francesco Fiorentino, metà del 1400), successivamente sottoposto a esame con interferometria olografica per rivelare la presenza di distacchi fra gli strati superficiali del dipinto e il supporto ligneo (*si vedano le figure a pag. 15*).

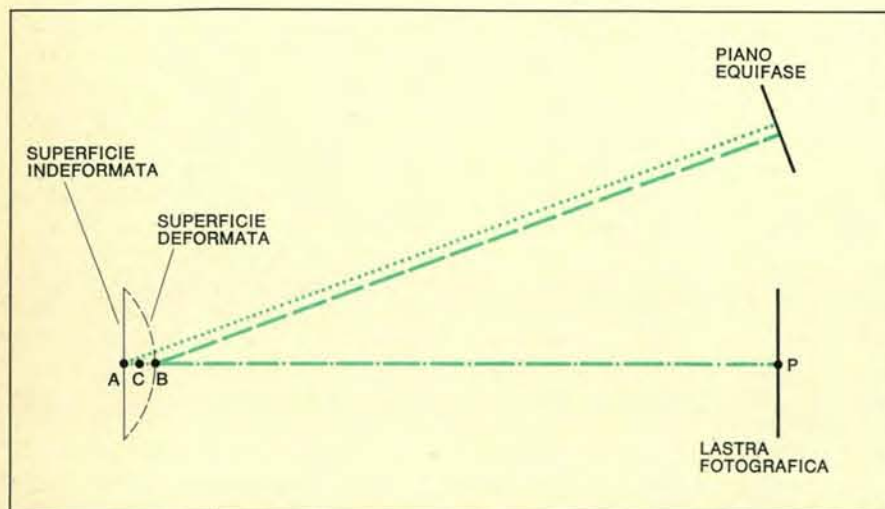
Non essendo ovviamente pensabile di correggere questo stato di cose abbandonando il criterio del rispetto dell'autenticità, viene quindi spontaneo chiedersi se il problema della conservazione delle opere d'arte ammetta soluzione. Per rispondere a questa domanda si può osservare che, da un punto di vista meramente tecnico, un'opera d'arte è pensabile come un messaggio (nel senso che a questo termine dà la teoria dell'informazione), la cui informazione è portata da elementi configurazionali e cromatici impressi sul supporto che costituisce materialmente l'opera stessa. Il deterioramento di tale supporto conduce a una progressiva perdita (ma forse sarebbe più esatto dire trasformazione) del contenuto di informazione.

In questo ordine di idee, la soluzione di tipo tecnico del problema è chiaramente la registrazione, in forma opportunamente codificata, del contenuto di informazione dell'opera su un supporto non labile e la sua successiva ritraduzione nel linguaggio originario. Gioverà alla chiarezza del nostro discorso l'osservare che un siffatto procedimento di registrazione e ritraduzione dell'informazione è già tradizionalmente usato in campo artistico. Stesura ed esecuzione di uno spartito sono, in campo musicale, i due momenti (registrazione, traduzione) del processo sopra detto.

Nel campo delle arti figurative, il mezzo tecnico cui è più spontaneo pensare per realizzare il procedimento sopra illustrato è la fotografia, i meriti e i limiti della quale sono così noti da non richiedere discussione. Più completo, dal punto di vista della registrazione, è il metodo fotogrammetrico che si presenta pertanto prezioso per tale fine (si veda l'articolo *Nuovi orizzonti nel rilievo e nel controllo dei monumenti* di Mario Fondelli e Francesco Greco in « Le Scienze », n. 63, novembre 1973), ma per il quale è di difficile realizzazione la fase di ritraduzione nel linguaggio originario. Il metodo più promettente, almeno in un certo numero di casi, è l'olografia (si veda l'articolo *Progressi in olografia* di Keith S. Pennington in « Le Scienze », n. 1,

settembre 1968). In esso, la fase di ritraduzione è, per il principio stesso del metodo, così completa e così fedele da non permettere all'osservatore di distinguere l'immagine ricostruita da quella originale. Va detto tuttavia che, sebbene alcuni esperimenti preliminari in questa direzione siano già stati condotti, l'applicazione di questo metodo è ancora tutta da studiare e che un certo numero di problemi tecnici (come quello della produzione di efficienti ologrammi a colori) vanno ancora risolti.

La soluzione al problema della conservazione ora prospettata, oltre a non essere immediatamente attuabile, cessa di essere valida se non si accetta la definizione di opera d'arte data in termini di contenuto d'informazione. Il problema può quindi ripresentarsi nella sua interezza. Noi ci limiteremo a osservare che è forse possibile operare sulle strutture e sui materiali antichi in maniera abbastanza raffinata da conoscere i meccanismi di deterioramento e tentare di rallentarli al massimo, senza turbare minimamente ciò che fa di questi materiali e strutture un'opera d'arte. Dal punto di vista scientifico è un obiettivo che ben si concilia con l'attuale tendenza delle ricerche a sviluppare in ogni campo metodi d'indagine non distruttivi e di buona ripetibilità, applicabili a oggetti dalla struttura complessa e dei quali importa conoscere, come per le opere d'arte, la risposta a sollecitazioni minime e di durata sia istantanea che lunghissima.

Schema di un apparato di registrazione per interferometria olografica. Il fascio di raggi paralleli emergenti dal laser (*in colore*) è suddiviso in due fasci; di questi uno (detto di riferimento) va a incidere, attraverso una lente divergente, sulla lastra fotografica; l'altro, opportunamente espanso da un sistema telescopico di lenti convergenti, serve a illuminare la superficie dell'oggetto. Si possono registrare sulla medesima lastra due ologrammi sovrapposti della superficie dell'oggetto uno nella posizione indeformata (*linea continua*) e uno nella posizione deformata (*linea a tratti*).

Un punto *A* della superficie dell'oggetto si porta, dopo la deformazione, nel punto *B*. Per i piccolissimi spostamenti studiati con l'interferometria olografica, quando la superficie indeformata e quella deformata vengono ricostruite simultaneamente dall'ologramma, i punti *A* e *B* appaiono nella lastra fotografica coincidenti in un unico punto *C*, la cui luminosità dipende dalla differenza dei cammini ottici dei raggi che illuminavano il punto *A* (*linea in colore punteggiata*) e il punto *B* (*linea a tratti*).

L'interferometria olografica, e in particolare quella versione di essa che prende il nome di olografia a doppia esposizione, è uno di questi metodi; l'applicazione che ne è stata fatta, e di cui riferiamo in questo articolo, è frutto d'una iniziativa presa congiuntamente dall'ENI-Servizio pianificazione tecnologica, dall'Istituto centrale del restauro di Roma e dall'Istituto di fisica della facoltà d'ingegneria dell'Università dell'Aquila.

Con l'interferometria olografica è possibile studiare le deformazioni subite da un oggetto di forma e natura qualsiasi, semplicemente registrando sulla stessa lastra due ologrammi di questo oggetto prima e dopo che le deformazioni si siano verificate. Quando, dopo sviluppo e fissaggio, si riillumina la lastra, vengono ricostruite due immagini corrispondenti alle configurazioni assunte dall'oggetto prima e dopo la deformazione. I campi ricostruiti sono coerenti e danno quindi luogo a fenomeni di interferenza. Per chi osservi l'oggetto attraverso l'ologramma, il fenomeno si manifesta nel fatto che esso appare solcato da strisce scure.

Può accadere che, spostando l'occhio da un punto all'altro dell'ologramma, e cambiando quindi l'angolo sotto cui si osserva l'oggetto, si vedano le frange scure scorrere su di esso; cioè può accadere che punti dell'oggetto inizialmente scuri (perché coperti da una frangia scura) diventino chiari quando si cambi il punto di osservazione. Viceversa può anche accadere che le frange non cambino posizione rispetto all'oggetto quando si cambi il punto di osservazione. Il verificarsi dell'una o dell'altra circostanza dipende dalla posizione in cui si formano nello spazio le frange di interferenza. In certi casi esse si formano in punti dello spazio che non appartegono alla superficie dell'oggetto, ad esempio in un piano intermedio fra l'oggetto e l'osservatore. È questo il caso in cui, spostando l'occhio attraverso la lastra, si vedono, per una sorta di fenomeno di parallasse, le frange scorrere sull'oggetto. Si dice allora che le frange non sono localizzate sull'oggetto. In altri casi invece le frange si formano sulla superficie dell'oggetto, e quindi non cambiano posizione rispetto a esso quando si cambia il punto di vista. Si parla allora di frange localizzate sull'oggetto.

La localizzazione delle frange dipende da diversi fattori, quali la configurazione spaziale del fascio con cui si illumina l'oggetto, il tipo di deformazione e la direzione media sotto cui si vede l'oggetto dall'ologramma. Il problema generale dell'interferometria olografica è quello di individuare lo spostamento di ciascun punto dell'oggetto a partire dall'esame delle frange di interferenza. Si tratta di un problema che, in generale, presenta notevoli difficoltà pratiche anche se, in linea teorica, è perfettamente risolvibile. Tuttavia, per le applicazioni che ci interessano possiamo limitarci a una classe di fenomeni nei quali l'interpretazione delle frange è piuttosto immediata.

Cercheremo di spiegare come vanno le cose in base a un modello semplificato. Consideriamo la situazione illustrata nella figura in alto a pagina 12. La superficie di un oggetto, dapprima approssimativamente piana, subisce una deformazione che la porta nella configurazione disegnata in tratteggio. In entrambi i casi la superficie è illuminata con un fascio di raggi paralleli (in colore). Su una stessa lastra si registrano due ologrammi dell'oggetto, prima e dopo che sia avvenuta la deformazione. Quando si ricostruisce il doppio ologramma così ottenuto, l'oggetto viene fatto rivivere simultaneamente nelle due configurazioni. Ciò che si osserva da un generico punto della lastra è la sovrapposizione dei campi di radiazione ottica diffusi dall'oggetto nelle due situazioni.

Fissiamo ora la nostra attenzione su un determinato punto dell'oggetto che, dalla posizione contrassegnata con *A*, si sia portato, dopo la deformazione, in quella contrassegnata con *B* (*si veda la figura in basso a pagina 12*). Guardiamo dal punto *P* dell'ologramma in direzione di *A* e *B*. Supporremo che la distanza fra *A* e *B* sia così piccola che, guardando dall'ologramma, i due punti non si vedano come distinti. È effettivamente questo il caso che ci interessa, essendo i comuni valori di tale distanza dell'ordine di alcuni micron. Possiamo quindi pensare che, al posto di *A* e *B*, si veda un solo punto *C* a metà strada fra i due. Per capire quale sarà la luminosità del punto *C* possiamo ragionare nel modo seguente. Tutto va come se nel punto *P* dell'ologramma giungessero le onde diffuse da *A* e da *B*. Nel sovrapporsi, queste due onde potranno rinforzarsi o indebolirsi a seconda della differenza di fase con cui giungono in *P*. Ad esempio, se la differenza di fase è un multiplo dispari di π, il che è quanto dire che istante per istante i campi elettrici delle due onde hanno verso opposto, le due onde si cancellano a vicenda, cosicché è come se dal punto *C* non venisse diffusa luce. Pertanto *C* sarà visto come un punto oscuro. Al contrario, se la differenza di fase è un multiplo pari di π (vettori elettrici concordi), l'ampiezza del campo risultante è doppia di quella di ciascun componente e l'intensità (che è proporzionale al quadrato del campo) si quadruplica. *C* sarà allora visto come un punto fortemente luminoso.

La differenza di fase, una volta fissata la direzione di osservazione e quella del fascio illuminante, dipende dalla distanza fra *A* e *B*. Per calcolarla si deve tenere presente che la radiazione, nell'andare dalla sorgente alla lastra attraverso la diffusione in *A* o in *B*, ha

Figura di interferenza ottenuta tramite ologramma a doppia esposizione di una tela che viene deformata. Fra la prima e la seconda esposizione, è stata esercitata un leggerissima pressione sul retro della tela, nel punto centrale. Nella figura di interferenza le linee di intensità costante possono considerarsi, nei riguardi della deformazione, analoghe a linee di livello. L'entità della deformazione aumenta dal bordo al centro, crescendo di mezza lunghezza d'onda ogni volta che si passa da una frangia luminosa alla successiva.

Figura di interferenza ottenuta con l'ologramma a doppia esposizione di una tavola da dipinto, durante un processo di riscaldamento. Le frange mostrano delle irregolarità in corrispondenza alle zone di passaggio fra fasci di fibre di legno adiacenti.

compiuto percorsi di diversa lunghezza a seconda che il centro diffusore sia A o B. La differenza fra i due percorsi, rapportata alla lunghezza d'onda della radiazione e moltiplicata per 2π, fornisce la differenza di fase cercata. Basta allora valutare i percorsi complessivi dei raggi luminosi che, attraverso A o B, giungono in P, a partire da un piano (equifase) ortogonale alla direzione dei raggi del fascio che illumina l'oggetto.

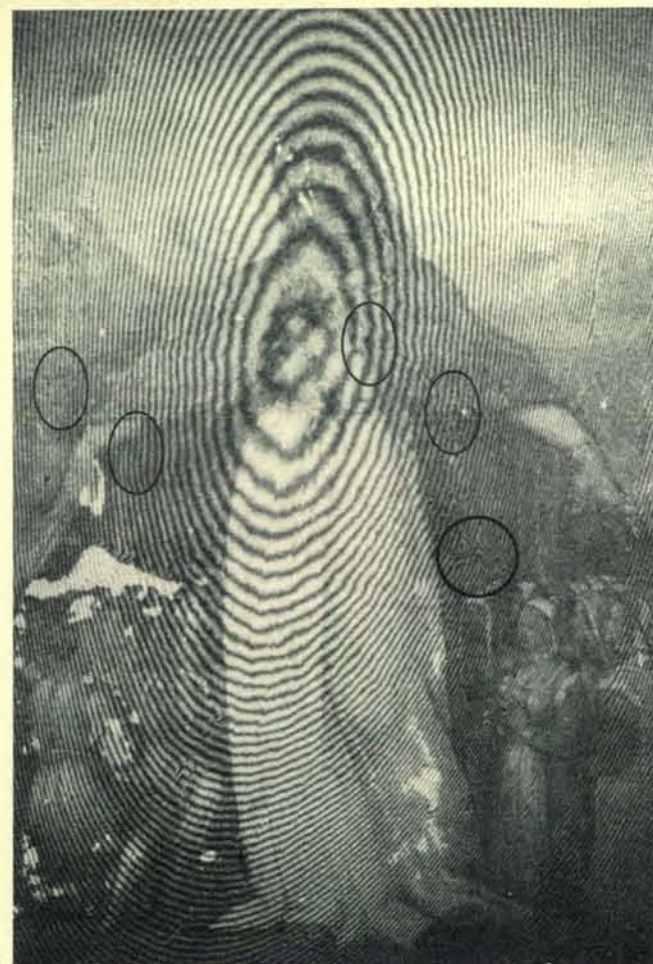
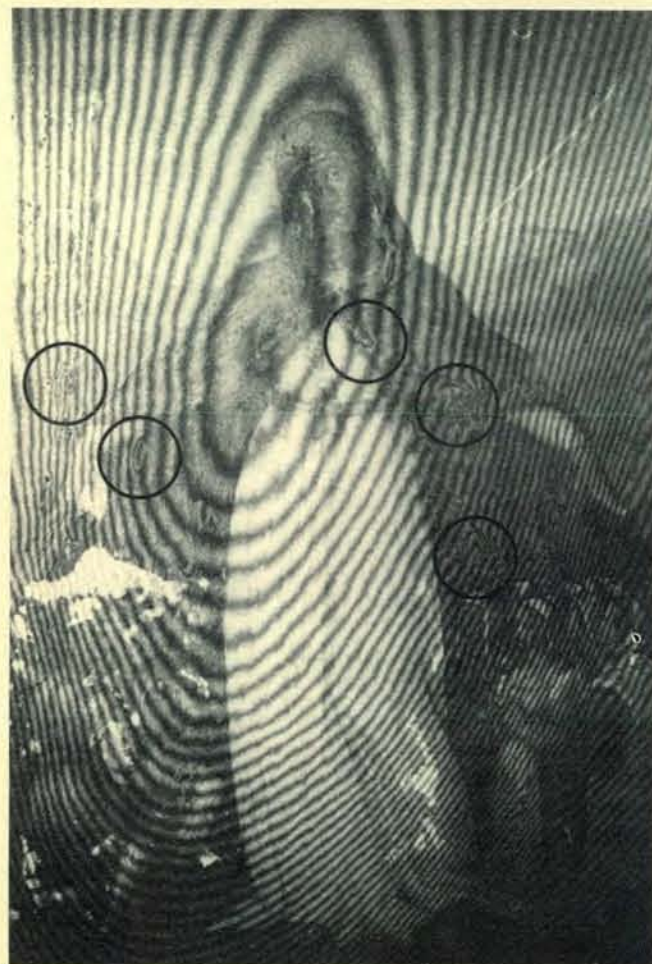
Dopo questa breve discussione del metodo dell'interferometria olografica, passiamo a illustrare qualche applicazione di esso a problemi connessi con il restauro. Innanzitutto possiamo distinguere due categorie di problemi: a) quelli relativi allo studio del comportamento di opere d'arte rispetto a sollecitazioni di vario tipo; b) quelli relativi alla diagnosi dello stato di conservazione di un'opera d'arte.

Cominciamo dalla prima categoria. Un'opera d'arte è normalmente soggetta a sollecitazioni di vario tipo: fisico, chimico, biologico. Variazioni di umidità e temperatura, agenti chimici presenti nell'atmosfera, effetti fotochimici, attacchi da microorganismi sottopongono l'opera d'arte a un processo di graduale modifica dei caratteri originari che può portare allo snaturamento totale di essa. È quindi essenziale, ai fini conservativi, poter determinare l'influsso di tali fattori sull'opera d'arte. Volta a volta, in dipendenza dai materiali costituenti l'opera e dalle caratteristiche dell'ambiente nel quale essa è conservata, l'uno o l'altro di tali fattori può assumere peso determinante. Le conoscenze che si hanno al riguardo sono piuttosto soddisfacenti per opere costituite da materiali di tipo semplice e omogeneo come pietra, metalli ecc. Sono invece più incomplete riguardo a opere di struttura e composizione eterogenea, quali dipinti su tela, su tavola ecc. In casi del genere ci si rifà spesso a dei modelli del sistema in studio. Intendiamo per modello una schematizzazione dell'oggetto reale che sia in grado di rendere conto di un certo numero di aspetti del suo comportamento. Prescindendo dalla difficoltà di messa a punto di un modello, vogliamo qui mettere in risalto il fatto che, in molti casi, è difficile sottoporre a verifica la validità del modello assunto mediante confronto con il comportamento dell'oggetto reale. Un esempio tipico è quello della verifica di validità di un modello che debba rendere conto dell'insieme delle deformazioni subite da un oggetto a seguito di sollecitazioni di tipo termico, meccanico ecc. Mentre è relativamente semplice misurare parametri globali (quali, ad esempio, allungamenti per trazione), si hanno serie difficoltà di ordine pratico e concettuale quando si voglia determinare una mappa completa di spostamenti puntuali. Taluni metodi, come ad esempio l'uso di sensori (*strain gauges*), pur essendo di indubbia utilità, soffrono di alcune limitazioni non soltanto riguardo al numero di informazioni locali che permettono di ottenere, ma anche in relazione alle modifiche di comportamento che, con la loro stessa presenza, essi possono indurre nella struttura e che, in taluni casi, possono essere tutt'altro che trascurabili. È in casi come questo che l'uso dell'interferometria olografica può essere di valido aiuto e presentare notevoli vantaggi rispetto ai metodi tradizionali. Fra questi vantaggi citiamo la sensibilità della misura, la possibilità di ottenere una mappa simultanea di tutta la superficie osservata dell'oggetto e la riduzione a livelli trascurabili della perturbazione indotta nell'oggetto dal processo stesso di misura (la perturbazione si riduce infatti alla semplice illuminazione dell'oggetto).

Per dare un esempio di applicazione a un problema specifico, ci riferiamo al caso seguente. È noto che uno degli interventi conservativi attuati nei dipinti su tela è la cosiddetta «rifoderatura». Tale operazione consiste fondamentalmente nel distaccare dal telaio la tela che porta il dipinto e nell'incollarla su una nuova tela, la quale, tesa su un nuovo telaio, sostituisce la vecchia tela nella funzione di supporto del dipinto. È chiaro che, per poter giudicare l'effetto di tale operazione sulla vita del dipinto, occorre quantomeno conoscere il comportamento di una tela da rifodero rispetto alle sollecitazioni ambientali. Anche qui si pone innanzitutto il problema della scelta di un modello. Essendo, allo stato attuale delle conoscenze, non certo facilmente ideabile un modello che sia in grado di rendere conto del comportamento della tela rispetto a tutte le possibili sollecitazioni a cui può essere sottoposta (quale che sia l'intervallo di tempo durante il quale si sviluppa il processo), sarà necessario rifarsi a dei modelli che tengano conto soltanto di un ben preciso tipo di sollecitazione e di una ben precisa scala temporale dei fenomeni. Ad esempio, per sollecitazioni meccaniche di modesta entità agenti per tempi brevi, si potrebbe adottare per la tela montata su telaio un semplicistico modello che la assimili a una membrana elastica fissata lungo il bordo. Se si adotta un tale modello, e se si suppone che la sollecitazione agisca al centro della membrana e perpendicolarmente al suo piano, è abbastanza semplice calcolare teoricamente lo spostamento normale subito punto per punto dalla tela. Si tratterà infatti di risolvere l'equazione di Laplace, con opportune condizioni al contorno che schematizzino l'azione dei vincoli perimetrali e lo spostamento imposto nella regione centrale. Questo è un tipo di problema che si incontra in molte branche della fisica matematica (elettrostatica, teoria dell'elasticità ecc.), e per il quale esistono tecniche di soluzione sia analitiche sia numeriche. Senza entrare nei dettagli matematici, sarà sufficiente rilevare che la soluzione dipenderà dallo stato di tensionamento della membrana, assumendo aspetti diversi a seconda che il tensionamento sia di tipo uniforme, o maggiore in una delle due direzioni parallele ai lati, o di tipo più complesso. A prescindere dalla validità pratica del suddetto modello, ciò che qui interessa porre in rilievo è come il confronto fra previsioni teoriche e comportamento reale possa farsi in maniera semplice e completa sfruttando l'interferometria olografica.

Infatti, mediante il procedimento prima illustrato, è possibile ottenere una mappa degli spostamenti effettivi della tela da confrontare con le mappe teoriche (*si veda la figura a pagina 13*) e da tale confronto possono ricavarsi diverse informazioni. In primo luogo si può controllare l'aderenza fra il modello teorico e l'oggetto reale; in secondo luogo si può individuare, nei limiti di validità del modello, la situazione di tensionamento della tela. Ad esempio, se si confrontassero le curve ottenute con quelle teoriche ricavabili risolvendo l'equazione di Laplace, si dedurrebbe facilmente che il modello a membrana rende conto bene del comportamento della tela su gran parte della superficie, a eccezione di una ristretta regione in vicinanza dei punti di applicazione della sollecitazione, dove, come era da attendersi, l'ipotesi di membrana elastica (cioè di rigidezza a flessione uguale a zero) comporta scostamenti dalla realtà. Così ancora è del

Frange di interferenza da ologramma a doppia esposizione del dipinto di pagina 10, ottenute sfruttando le deformazioni del dipinto conseguenti alle normali variazioni dei parametri ambientali (umidità, temperatura). L'irregolarità delle frange in alcuni punti indica la presenza di zone di distacco della pittura.

Frange di interferenza ottenute con ologramma a doppia esposizione del dipinto di pagina 10 registrato durante un processo di riscaldamento della superficie del dipinto, mediante aria calda. Le zone di distacco, messe in evidenza dalle irregolarità delle frange, sono molto più numerose che nella figura accanto.

Esame di un dipinto («Madonna della Misericordia», scuola del Perugino) mediante ologramma a doppia esposizione. Alcune zone di distacco sono state indicate entro i cerchietti nella figura a sinistra. La figura a destra è stata eseguita dopo che su alcune delle zone distaccate (ora contrassegnate con delle ellissi) è stato operato un intervento di consolidamento. Si noti la regolarità delle frange dove è stato operato il consolidamento (nelle ellissi) a confronto dell'area non consolidata.

tutto immediato dedurre per esempio che lo stato di tensionamento lungo i due assi della tela può essere differente in condizioni diverse e così via. Se informazioni di questo tipo possono addirittura ricavarsi con una occhiata, il confronto fra i rilievi quantitativi delle deformazioni e i dati teorici permetterà di avere un'informazione dettagliata e quantitativa sullo stato di tensione della tela.

Vogliamo ora passare a illustrare l'applicazione dell'interferometria olografica alla seconda categoria di problemi nel campo del restauro, e precisamente il suo uso come metodo di diagnosi di danni allo stato iniziale o comunque di difficile rilevamento. Ci riferiamo principalmente ai dipinti su tavola, sebbene molte delle considerazioni che stiamo per svolgere possono estendersi ad altre categorie di opere d'arte. Cominciamo con il ricordare che un dipinto su tavola (come del resto altri tipi di dipinti) può essere considerato come una struttura stratificata. Difatti il supporto ligneo (generalmente pioppo nella pittura italiana, rovere in quella fiamminga) è ricoperto da una serie di strati di preparazione in colla e gesso, al di sopra dei quali è disteso lo strato dei colori, cioè il dipinto propriamente detto. Durante la vita del dipinto è frequente che all'interno di questa struttura si verifichino dei difetti di adesione. Essi possono essere localizzati a diverse profondità entro la stratificazione, ma il caso più frequente è dato dai distacchi nell'interfaccia fra il supporto ligneo e gli strati di preparazione. A seguito della presenza di tali distacchi, le regioni stratificate a essi sovrastanti si comportano rispetto alle sollecitazioni ambientali in modo differente dalle restanti regioni (grossolanamente parlando con deformazioni maggiori). Da un lato ciò tende a far aumentare l'area della regione distaccata, dall'altro fa sì che gli strati sovrastanti le regioni di distacco tendano a sollevarsi rispetto al resto del dipinto, formando rigonfiamenti che, successivamente, danno luogo alla rottura della zona di confine fra strati aderenti e non. Esiste pertanto una fase del processo nella quale gli strati sono distaccati dal supporto, ma non si è ancora verificata la rottura. È in questa fase, la cui durata è generalmente tanto maggiore quanto maggiore è lo spessore degli strati distaccati, che è possibile l'intervento di restauro mediante iniezioni di consolidamento che ripristinino l'adesione fra gli strati e il supporto. È quindi di grande importanza il disporre di metodi di diagnosi che permettano di individuare la presenza di distacchi di questo tipo. La diagnosi sarà poi tanto più utile quanto maggiore è la sensibilità del metodo, cioè quanto più piccolo è il difetto che può essere rilevato. Difatti, per quanto manchino dati sistematici in proposito, si può affermare con sicurezza che la probabilità di successo dell'intervento di restauro valutato su tempi lunghi sarà tanto più alta quanto minore è la perturbazione strutturale introdotta dall'intervento stesso e quindi quanto minore è l'estensione della regione danneggiata, dato che l'operazione di consolidamento comporta l'iniezione di materiali di natura necessariamente diversa da quella degli strati originari.

Il metodo di diagnosi che ora illustreremo si basa sull'interferometria olografica e si fonda sul fatto, già notato sopra, che le regioni distaccate si comportano in maniera differente da quelle non distaccate, quando il dipinto sia assoggettato a sollecitazioni. L'esame interferometrico delle deformazioni del dipinto metterà allora in evidenza le regioni di distacco come quelle nelle quali si notano delle anomalie nell'andamento delle frange rispetto all'andamento medio.

Per sviluppare un programma di questo tipo è necessario assoggettare il dipinto a una sollecitazione che provochi una deformazione. Peraltro, potendo essere assai piccole le deformazioni che il metodo interferometrico è in grado di rivelare, è chiaro che la sollecitazione imposta al dipinto può essere così debole da non comportare alcun rischio per esso.

La sollecitazione più semplice da applicare è quella derivante da variazioni dei parametri ambientali. A questo proposito è opportuno rilevare che i processi di adeguamento di una tavola lignea a variazioni di umidità e temperatura sono molto lenti (le loro durate si misurano sulla scala dei giorni). Ne segue che un dipinto su tavola, posto in un ambiente nel quale non vi siano apparati di stabilizzazione dell'umidità e della temperatura, è soggetto a un continuo processo di adattamento derivante dalle variazioni dei parametri ambientali. Si possono allora sfruttare le deformazioni che accompagnano tali processi per mettere in evidenza i difetti. Un intervallo di tempo di alcuni minuti è generalmente sufficiente a produrre deformazioni rivelabili interferometricamente. Un metodo alternativo consiste nel riscaldare debolmente la superficie del dipinto mediante irraggiamento o mediante un getto di aria calda. C'è da attendersi che questo metodo, agendo direttamente sulla superficie, sia più adatto del precedente a evidenziare i difetti degli strati superficiali.

Per poter giudicare l'efficacia di questi metodi dobbiamo innanzitutto chiederci quale figura di interferenza è da attendersi nel caso di un dipinto che non abbia distacchi. Riferiamoci al caso in cui il dipinto venga riscaldato. Negli esempi che riporteremo i dipinti erano vincolati alle estremità della mediana maggiore mediante due tasselli di legno incollati sul retro della tavola. In queste condizioni il movimento principale sarà un incurvamento intorno alla mediana maggiore. La componente principale del sistema di frange sarà pertanto costituita da frange parallele alla mediana maggiore, che si infittiscono dal centro verso i bordi. Una deformazione maggiore del centro della tavola (non direttamente vincolato), rispetto alle estremità della mediana maggiore, potrà far degenerare queste frange rettilinee in curve chiuse, molto grossolanamente approssimabili a ellissi. Nel caso di deformazione derivante da variazioni dei parametri ambientali, la situazione potrà risultare più complessa, ma la componente principale del sistema di frange sarà la stessa. Queste valutazioni possono essere controllate operando su tavole lignee non dipinte che siano preparate *ad hoc*. Bisogna qui distinguere il caso in cui la tavola sia nuda, da quello in cui una delle sue facce sia stata ricoperta con gli strati di preparazione in colla e gesso. Nel primo caso si rileva l'effetto delle fibre del legno sulla figura di interferenza, effetto consistente in una deformazione locale delle frange ogni volta che si attraversa il margine fra un fascio di fibre e il successivo (*si veda la figura a pagina 14*). Tale effetto viene mediato dalla presenza degli strati di preparazione.

Una volta stabilito l'andamento di base delle frange si può sottoporre a prova il metodo di diagnosi. Presenteremo a tale scopo, quale esempio, i risultati ottenuti su un dipinto della metà del 1400, la «Santa Caterina» di Pier Francesco Fiorentino, quale si presenta attraverso un ologramma ordinario (*si veda la figura a pagina 10*), e i sistemi di frange ottenute sfruttando i fenomeni di deriva ambientale e quelli ottenuti con riscaldamento della superficie. Come si vede dalle figure di pagina 15, la presenza di distacchi viene evidenziata dallo scostamento delle frange dall'andamento medio; da notare che il metodo che sfrutta il processo di riscaldamento della superficie è notevolmente più sensibile di quello che si basa sulla deriva ambientale.

Per quanto riguarda la sensibilità del metodo va detto che, mentre nel caso del dipinto raffigurante S. Caterina un certo numero di distacchi aveva raggiunto entità tale da permetterne l'individuazione anche con i metodi tradizionali, nel caso di altri dipinti gli esami tradizionali non mettevano in evidenza alcun distacco. Per esempio in un dipinto raffigurante una Madonna della Misericordia della scuola del Perugino se si osservano le regioni del dipinto contrassegnate da un circoletto (*si veda la figura a pagina 16 a sinistra*), si vede che in esse le frange si scostano dall'andamento medio rivelando la presenza di distacchi. Essendo tali distacchi, come si è detto, di entità così ridotta da non poter essere rivelati con i metodi tradizionali, su alcuni di essi fu operato un intervento di consolidamento al fine di verificare l'esattezza della diagnosi. Il successivo esame olografico mostrò (*si veda la figura pagina 16 a destra*) che le anomalie scomparivano nelle regioni restaurate, mentre permanevano in quelle non restaurate. Ciò è evidenziato nella figura citata dove le zone restaurate sono contrassegnate da un'ellisse.

Il metodo ora illustrato può essere esteso anche a statue policrome, ma in questo caso l'interpretazione dei sistemi di frange ottenuti risulta più complessa a causa della diversa giacitura delle varie porzioni di superficie dell'opera. Anche se gli scostamenti locali dovuti a difetti sono in questo caso di individuazione meno immediata, il metodo rimane valido.

Da questi primi esperimenti si può concludere che l'interferometria olografica si è rivelata come un metodo d'indagine di grande precisione per chiarire il comportamento meccanico dei dipinti, sia su tela che su tavola, e delle sculture policrome sotto l'azione di fattori ambientali quali l'umidità e la temperatura. È dunque ora possibile, per i direttori dei musei e per i restauratori, conoscere fin nei più minuti dettagli la complessa vicenda di sforzi e deformazioni che ha luogo, si può dire senza un attimo d'interruzione, sui materiali e sulle strutture che costituiscono le opere d'arte affidate alle loro cure. Questo tipo di conoscenza è certamente molto utile per determinare le condizioni ambientali che, producendo il minimo di sforzi e deformazioni, risultino le più favorevoli alla conservazione delle opere d'arte.

Inoltre, gli ologrammi a doppia esposizione mettono in evidenza difetti strutturali, quali la mancanza d'adesione tra vari strati di colore, di cui non era finora possibile accertare la presenza se non troppo tardi, cioè solo quando si fossero concretati come veri e propri danni. È evidente che quanto più è precoce la diagnosi, cioè quanto è meno grave il danno da riparare, tanto è più efficace e sicuro l'intervento riparativo. Non bisogna infine dimenticare che interventi del genere sono finora stati praticati per così dire alla cieca, nel senso che non era possibile controllare oggettivamente se e in che misura avessero effetto al fine specifico. Questo controllo può ora essere fatto ripetendo l'ologramma a doppia esposizione dopo l'intervento, ed eventualmente replicandolo, come un *check-up*, a scadenze prefissate.

Energia di fusione da implosione con il laser

Il principio della fusione nucleare mediante il laser è basato sull'ignizione di una pastiglia di combustibile fatta implodere da fasci laser focalizzati e compressa fino a densità altissime

di John L. Emmett, John Nuckolls e Lowell Wood

Il primo laser fu fatto funzionare nel 1960 da Theodore H. Maiman. Meno di un anno più tardi Stirling A. Colgate, Ray E. Kidder, John Nuckolls, Ronald Zabawski e Edward Teller allestirono un programma di calcoli mediante un elaboratore per cercare di scoprire che cosa accadeva quando piccolissime pastiglie di deuterio-trizio venivano portate alle condizioni termonucleari mediante implosione ottenuta con intensi fasci di luce laser. Si doveva anche vagliare la possibilità che le microesplosioni di fusione fossero utilizzabili per la produzione di energia. Quei primi calcoli mostrarono che per un'efficiente produzione di energia di fusione non bastava riscaldare col laser il combustibile termonucleare, ma era necessario comprimerlo con il laser fino a ottenere una densità 10 000 volte superiore a quella del suo stato liquido normale. Anche se la fusione col laser fu oggetto di frequenti studi già verso la fine degli anni '60, le idee fondamentali (tra cui l'implosione) furono considerate segreto militare dalla Commissione americana per l'energia atomica fino al 1972.

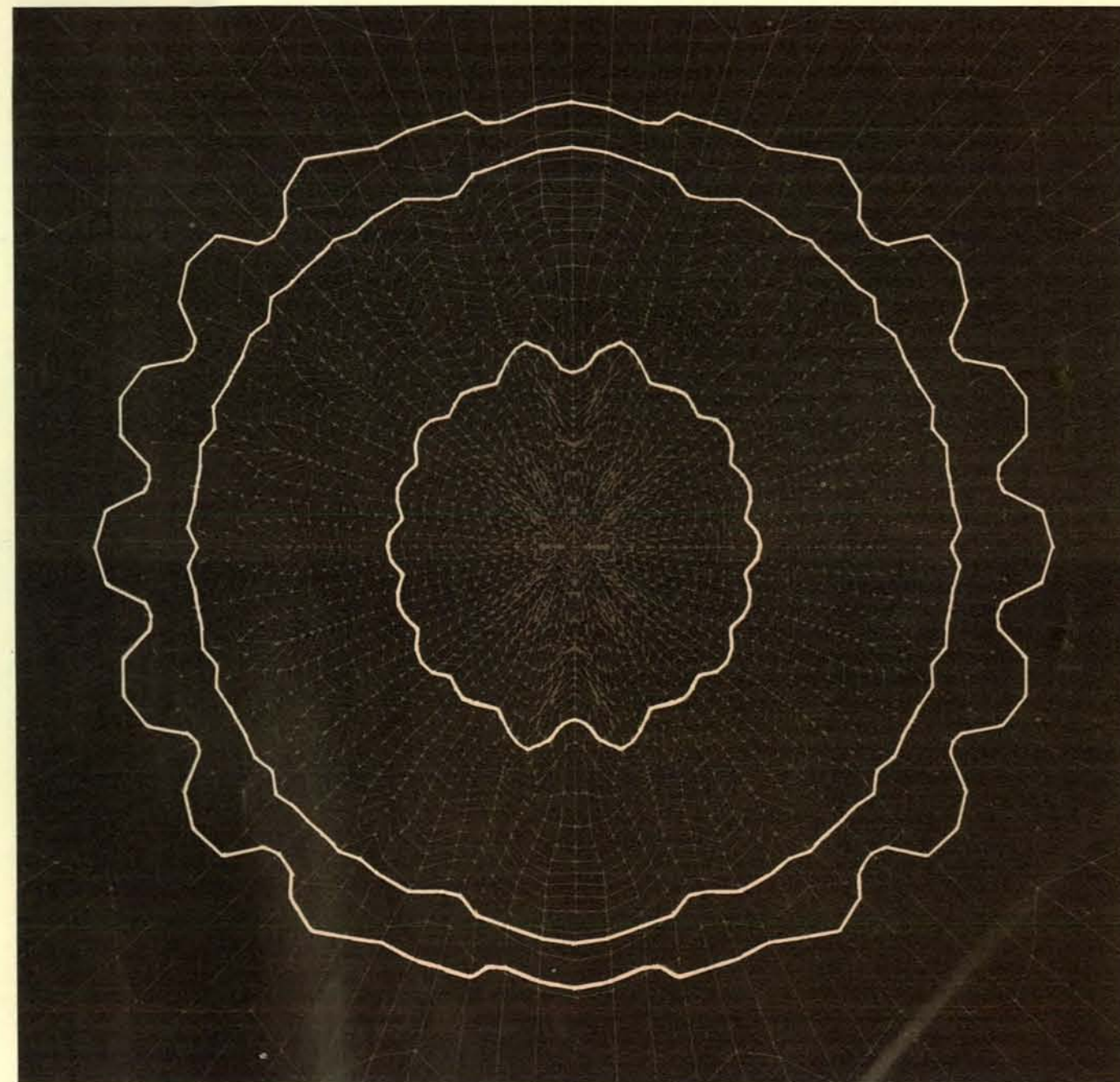
Modesti programmi sperimentali presero il via sin dall'inizio degli anni '60 al laboratorio Lawrence di Livermore della Commissione e successivamente in altri laboratori del mondo. Tuttavia gli esperimenti più indicativi non poterono essere eseguiti poiché con un breve impulso laser non si poteva fornire un'energia sufficiente. Con il rapido progredire della tecnologia laser, nel 1968 N.G. Basov e i suoi collaboratori dell'istituto di fisica Lebedev nell'URSS furono in grado di annunciare l'osservazione dei primi neutroni emessi da un plasma riscaldato col laser.

Oggi il nostro laboratorio di Livermore, il laboratorio scientifico di Los Alamos e diversi gruppi nell'URSS stanno progettando la realizzazione di laser in grado di fornire circa 10 000 joule di energia in impulsi della durata di un nanosecondo (un miliardesimo di secondo) o meno ancora. Con tali laser dovrebbe essere possibile eseguire decisivi esperimenti di fusione con il laser. Se si riuscirà a mettere a punto un efficiente sistema di ignizione del combustibile termonucleare, si potranno realizzare sistemi laser progrediti capaci di produrre energia su scala commerciale. I nostri studi attuali indicano che per un impianto di potenza da 1 gigawatt (un milione di kilowatt) sono necessarie 100 microesplosioni di fusione al secondo.

Il problema principale per la liberazione di energia termonucleare è il confinamento di un plasma caldissimo di nuclei leggeri, quali gli isotopi dell'idrogeno, per un tempo sufficientemente lungo perché possa avvenire una reazione. All'interno delle stelle plasmi di comune idrogeno e di altri elementi leggeri vengono portati alle temperature di ignizione e confinati dalla pressione gravitazionale. Nelle esplosioni termonucleari gli isotopi pesanti dell'idrogeno (deuterio e trizio) vengon riscaldati da un esplosivo a fissione. Il plasma prodotto viene confinato dalla sua stessa inerzia, cioè dal tempo finito necessario per accelerare e muovere del materiale fino a una distanza significativa. Il primo metodo per controllare la liberazione di energia termonucleare era basato sull'uso di bottiglie magnetiche per il confinamento di plasma di deuterio e trizio (DT). La fusione col laser riprende la tecnica del confinamento inerziale mediante la quale i laser riscaldano i plasmi con brevi impulsi di luce di densità di potenza estremamente alta. Negli ultimi quattro anni l'AEC ha più che duplicato lo stanziamento di fondi per il programma di fusione col laser, di modo che oggi l'impegno di ricerca per la fusione col laser è paragonabile a quello per il confinamento magnetico.

Perché un dato combustibile termonucleare possa bruciare efficientemente devono essere raggiunte certe condizioni indipendentemente dallo schema di confinamento adottato. Per prima cosa, si deve raggiungere la temperatura di ignizione termonucleare che per il DT è di 100 milioni (10^8) di gradi kelvin. Sono necessarie tali temperature per impartire ai nuclei velocità termiche sufficienti a far loro vincere la mutua repulsione elettrica che si verifica durante una collisione; a siffatte temperature la reazione nucleare procede in modo così rapido che l'energia prodotta è maggiore dell'energia dissipata per radiazione dal plasma. In secondo luogo, il plasma deve essere confinato abbastanza a lungo e con una densità abbastanza alta da consentire la reazione per una considerevole frazione dei nuclei. Questo secondo requisito è espresso dal numero di Lawson (ideato dal fisico inglese J.D. Lawson), definito come il prodotto della densità delle particelle interagenti per il tempo di confinamento. Il numero di Lawson, espresso in secondi per centimetro cubo, indica il raggiungimento del punto di «pareggio» (si veda l'articolo *Le prospettive dell'energia di fusione*, di William C. Gough e Bernard J. Eastlund, in «Le Scienze», n. 33, maggio 1971), basato sull'ipotesi che non più di un terzo dell'energia liberata debba essere utilizzata per sostenere la reazione e che la temperatura del plasma sia di 10^8 kelvin.

Il numero di Lawson è all'incirca lo stesso per la fusione col laser e per la fusione a confinamento magnetico: circa 10^{14} secondi per centimetro cubo per una miscela di combustibile DT. Nel confinamento magnetico l'obiettivo è la realizzazione di bottiglie magnetiche che siano in grado di confinare i combustibili per tempi relativamente lunghi; la massima densità viene determinata dal massimo campo magnetico che i materiali disponibili possono sopportare. Cifre attendibili sono un tempo di confinamento di circa un secondo e una densità di circa 10^{14} particelle per centimetro cubo. Nella fusione col laser l'obiettivo è di comprimere i combustibili a densità elevatissime (circa 10^{26} particelle per centimetro cubo); in tal caso il tempo efficace di confinamento è determinato dall'inerzia della materia (poco più di 10^{-12} secondi). Nei primi reattori progettati secondo l'uno o l'altro dei due meccanismi i neutroni liberati nelle reazioni termonucleari verranno assorbiti da uno schermo di litio, facendone aumentare la temperatura. Il calore asportato dallo schermo può essere impiegato per generare vapore e produrre elettricità nel modo tradizionale.

I principi della fusione a confinamento inerziale e la funzione dell'implosione possono essere facilmente chiariti con un semplice esempio. Supponiano di disporre di un laser in grado di produrre un milione di joule di ener-

L'implosione provocata dal laser di una pastiglia da 60 microgrammi degli isotopi dell'idrogeno deuterio e trizio viene simulata in forma bidimensionale da un programma di calcolo al laboratorio Lawrence di Livermore, dell'Università di California. Questo è un fotogramma del film prodotto dall'elaboratore elettronico. Il raggio iniziale della pastiglia prima di essere riscaldata e compressa da 11 fasci laser è di 400 micromètri (0,4 millimetri). Quando la pastiglia viene compressa fino a un raggio di circa 15 micromètri, il suo centro raggiunge la temperatura di ignizione, 10^8 kelvin. Nel fotogramma qui riprodotto, circa sette picosecondi (7×10^{-12} secondi) dopo la ignizione, la pastiglia si è riespansa fino a un raggio di circa 17 micromètri, corrispondente all'incirca alla posizione dell'anello bianco intermedio dei tre, che rappresentano isoterme della temperatura degli ioni. L'anello più interno rappresenta un fronte di detonazione termonucleare a 5×10^8 kelvin che si propaga attraverso la pastiglia dopo due fronti precedenti, uno a 2×10^8 kelvin (*anello centrale*) e uno a 10^8 kelvin (*anello esterno*). Il reticolo sullo sfondo rappresenta i moti idrodinamici. Il film è stato girato da George Zimmermann e Albert Thiessen.

gia ottica. Questa quantità di energia è appena sufficiente a riscaldare un milligrammo del combustibile nucleare di più rapida combustione (DT) alla temperatura di ignizione di 10^8 kelvin. Alla normale densità liquida un milligrammo di DT forma una sfera di circa un millimetro di raggio. L'efficienza con cui può essere bruciato il combustibile nel confinamento inerziale dipende dalla competitività tra il tempo di confinamento e il tempo di reazione di fusione.

Per una pastiglia di un millimetro alla temperatura di ignizione il tempo di confinamento inerziale (che è proporzionale al rapporto tra il raggio della pastiglia e la velocità termica dei nuclei) è di circa 2×10^{-10} secondi. Il tempo di reazione di fusione (che è inversamente proporzionale alla densità della pastiglia) è circa 1000 volte maggiore, cioè circa 2×10^7 secondi per un combustibile alla densità liquida. Di conseguenza solo un millesimo, cioè lo 0,1% della pastiglia brucerà, producendo solo un terzo del milione di joule del laser all'ingresso.

Se però la stessa pastiglia viene compressa di un fattore 10 nel raggio (cioè un fattore 1000 nel volume), il tempo di confinamento si riduce ugualmente di un fattore 10, mentre – ed è il fatto più importante – il tempo di combustione si riduce di un fattore 1000. Il risultato è che l'efficienza di combustione aumenta di un fattore 1000, fino al 10%, producendo un'energia pari a 30 volte quella laser in ingresso. È evidente che qualsiasi tentativo di produrre energia mediante fusione col laser basato sul riscaldamento di una pastiglia di combustibile termonucleare alla sua densità normale è destinato al fallimento. Per poter ottenere una liberazione di energia in quantità sensibilmente maggiore dell'ingresso laser, è indispensabile una compressione convergente della pastiglia di combustibile.

Ma non è tutto. Risulta che, se si comprime convenientemente una pastiglia, non è più necessario portare l'intera massa del combustibile alla temperatura di ignizione, e quindi diminuisce la quantità di energia laser necessaria. L'energia liberata nella fusione del DT viene trasportata principalmente da neutroni da 14 milioni di elettronvolt (MeV) e da particelle alfa (nuclei di elio) da 3,5 MeV. Se tali particelle vengono liberate in un plasma a 10^8 kelvin con densità pari a quella del DT liquido, esse hanno un cammino libero più di 10 volte maggiore del raggio della pastiglia da un millimetro, quindi quasi tutta l'energia sfugge senza riscaldare il resto della pastiglia. I cammini dei neutroni e delle particelle alfa sono però inversamente proporzionali alla densità del mezzo; di conseguenza, se essi vengono liberati in un plasma la cui densità è 1000 volte maggiore di quella liquida, il cammino di un neutrone risulta solo poche volte maggiore del raggio di 0,1 millimetri della pastiglia compressa, mentre quello di una particella alfa risulta minore di un sesto del raggio. Le particelle alfa cederanno quindi energia all'interno della pastiglia dando origine alla propagazione di un fronte di combustione termonucleare. Il risultato è che l'effettiva energia di ignizione viene ridotta di quasi un fattore 100. È una situazione davvero fortunata; diversamente la produzione di energia mediante fusione col laser non sarebbe possibile, poiché l'energia prodotta nella combustione del DT non è sufficientemente maggiore dell'energia di ignizione da compensare le perdite energetiche del laser, dell'implosione e del reattore (*si vedano le figure nella pagina a fianco e nella pagina 22*).

Una pastiglia di combustibile deuterio-trizio (DT) implode quando viene riscaldata simmetricamente da fasci laser focalizzati. Il massimo assorbimento di luce ha luogo in corrispondenza alla superficie critica, una regione ristretta nell'atmosfera a bassa densità che circonda la pastiglia. Gli elettroni « caldi » trasportano energia verso l'interno attraverso l'atmosfera restante per riscaldare la superficie della pastiglia e produrre ablazione. Il materiale così asportato esplode verso l'esterno mentre la reazione d'inerzia fa accelerare la pastiglia verso l'interno e la comprime fino a densità altissime.

L'inevitabile necessità di comprimere il combustibile termonucleare ad alte densità della fusione col laser basata sul confinamento inerziale ci fa porre la seguente domanda: le densità richieste sono almeno concettualmente raggiungibili? Tale possibilità è suggerita dalla sorprendente constatazione che l'energia necessaria per comprimere l'idrogeno fino a una densità pari a 10 000 volte quella liquida è solo l'1% dell'energia necessaria per riscaldare il DT alla temperatura di ignizione. Dal punto di vista energetico, perciò, la compressione è praticamente gratis rispetto all'ignizione. Per ottenere una compressione di un fattore 10 000, tuttavia, si devono produrre pressioni veramente astronomiche, qualcosa come 10^{12} atmosfere. Pressioni di questo ordine di grandezza esistono nel nucleo delle stelle, dove vengono generate con meccanismo gravitazionale da masse sovrastanti di 10^{33} grammi o più.

Si possono generare tali pressioni sulla Terra? Le nostre analisi indicano che implosioni sferiche generate col laser dovrebbero raggiungere o superare tali valori. Nell'implosione sferica generata col laser la luce viene focalizzata da una lente (o da uno specchio) su un'atmosfera a bassa densità di una certa sostanza, mentre questa evapora dalla superficie di una piccolissima pastiglia sferica (*si veda la figura a pagina 20*). L'intensa luce laser focalizzata viene assorbita nell'atmosfera di plasma a bassa densità mediante collisioni ione elettrone o mediante instabilità di plasma, producendo elettroni « caldi » con energie di alcune migliaia di elettronvolt. Quasi tutto l'assorbimento ha luogo in quella che è chiamata regione del plasma a densità critica, nella quale la frequenza delle oscillazioni di plasma uguaglia quella della luce laser.

La distanza che la luce laser deve percorrere prima di essere assorbita mediante collisioni nella regione a densità critica è proporzionale al quadrato della lunghezza d'onda della radiazione; così in un plasma a 10^7 kelvin la lunghezza d'onda di assorbimento per collisione è di un millimetro per radiazioni con lunghezza d'onda di un micromètro ed è di 100 millimetri per radiazioni con lunghezza d'onda 10 volte maggiore. Poiché le dimensioni della pastiglia sono di circa un millimetro, una radiazione da 10 micromètri non può essere assorbita efficacemente con un processo di collisione. Però, per fortuna, nelle implosioni col laser che ci interessano l'altissima densità della radiazione da 10 micromètri eccita instabilità di plasma che portano a un assorbimento efficace.

Gli elettroni caldi prodotti nei due processi diffondono nell'atmosfera lungo il gradiente termico radiale per riscaldare la superficie della pastiglia, che viene continuamente raffreddata per ablazione: è lo stesso meccanismo che raffredda lo scudo termico di un veicolo spaziale al rientro nell'atmosfera. Quando il materiale asportato schizza via, genera una forza uguale e opposta (secondo la terza legge di Newton) che produce l'implosione. In pratica il sistema è un razzo sferico che riceve energia dal laser a scapito di una rapida contrazione della pastiglia di combustibile. Il rendimento energetico dell'implosione del razzo è basso – minore del 10% – poiché la velocità di scappamento (del materiale asportato) è molto maggiore della velocità del veicolo (la pastiglia sotto implosione).

La materia compressa viene accelerata verso l'interno con una velocità quasi 50 volte maggiore della velocità di fuga dalla Terra e di conseguenza essa si collassa all'interno fin quando la pressione prodotta nella materia compressa frena l'implosione. L'effetto globale di questi processi è di aumentare la potenza per unità di superficie di 14 ordini di grandezza, da 10^5 watt per centimetro quadrato nel sistema a pompaggio laser a 10^{19} watt per centimetro quadrato nell'implosione (*si vedano le figure della pagina 23*). Contemporaneamente l'energia viene trasferita dai fotoni laser agli ioni di deuterio e trizio che innescano la reazione.

Se la pastiglia non viene compressa con molta cautela sono necessarie pressioni ben maggiori di 10^{12} atmosfere. È quindi necessario modellare accuratamente l'andamento della pressione in funzione del tempo dell'implosione a partire dall'andamento temporale dell'impulso laser. Per una pastiglia di DT solido la pressione attuale di innesco passa da 10^6 a 10^{11} atmosfere nell'intervallo di 10^{-8} secondi. In queste condizioni la velocità di implosione è sempre paragonabile alla velocità locale del suono nella pastiglia. Il

La compressione del combustibile necessaria per la fusione laser viene determinata dalle previsioni sui rendimenti del laser (10 %), dell'implosione (5 %) e della generazione di energia elettrica (in pratica il 10 %, poiché per il pompaggio del laser non dovrebbe essere usato più di un terzo dell'energia elettrica prodotta). Per provocare l'ignizione del DT sono necessari circa 10^9 joule per grammo e vengono liberati circa 10^{11} joule per grammo se fonde effettivamente il 30 % del combustibile. Il guadagno di energia di 100 non è sufficientemente grande da compensare i vari rendimenti elencati ($0,10 \times 0,05 \times 0,10$, uguale a 0,0005 cioè un rendimento globale dello 0,05 %). Per fortuna, in una pastiglia di combustibile fortemente compressa l'energia di fusione liberata da una piccola porzione di pastiglia è sufficiente per l'ignizione della restante parte mediante propagazione radiale, per cui l'energia laser richiesta in ingresso si riduce di circa un fattore 100. Così, per compressioni incluse nella gamma tra 10^3 e 10^4, la differenza tra l'energia laser minima all'ingresso (circa 2×10^7 joule per grammo per la compressione di DT relativamente freddo) e l'energia prodotta con un rendimento di combustione del 30 % è abbastanza grande da rendere economica la fusione nucleare.

risultato è una compressione della maggior parte della pastiglia senza apprezzabile riscaldamento, mentre la regione centrale viene deliberatamente portata verso la temperatura di ignizione.

È ben noto che i ricercatori che stanno cercando di ottenere la fusione controllata col metodo del confinamento magnetico si imbattono in vari tipi di instabilità del plasma che provocano perdite nelle bottiglie magnetiche. Ci si potrebbe domandare: il metodo laser è anch'esso caratterizzato da simili problemi? La risposta è positiva, ma le difficoltà appaiono superabili. Le instabilità del plasma possono in effetti essere eccitate quando vengono focalizzati laser di altissima potenza sull'atmosfera che circonda la pastiglia.

Il campo elettrico oscillante della luce laser accelera gli elettroni e gli ioni del plasma in verso opposto ed essi acquistano quindi una velocità di oscillazione relativa. Se tale velocità è sufficientemente alta, questo movimento provoca nel plasma onde instabili che crescono nel tempo, con effetto analogo a quello dei venti forti sulle onde del mare. Sebbene questa intensa turbolenza del plasma assorba efficacemente la luce laser, essa può anche produrre elettroni estremamente energetici (supertermici). Questi elettroni possono penetrare in profondità nella pastiglia, riscaldandola prematuramente e rendere in talmodo più difficile la compressione. Questo fenomeno è detto preriscaldamento. Gli elettroni supertermici, inoltre, possono avere un cammino talmente lungo che le collisioni con gli elettroni « più freddi » diventano meno frequenti, al punto che la velocità di scambio termico tra l'atmosfera e la pastiglia cala bruscamente e la pastiglia non subisce una compressione adeguata. Questo fenomeno è detto disaccoppiamento.

Allo scopo di evitare sia il preriscaldamento che il disaccoppiamento, i primi esperimenti di implosione col laser sono stati progettati con intensità laser tali da non superare le soglie di instabilità del plasma. Tali soglie possono essere alzate aggiungendo alla pastiglia tracce di materiali di alto numero atomico (per aumentare la frequenza delle collisioni) e usando laser con lunghezze d'onda minori e larghezze di banda maggiori. Un altro espediente potrebbe essere l'impiego di pastiglie cave, che consentono una più bassa pressione di ablazione per un tempo più lungo e su superfici e volumi maggiori rispetto alle pastiglie piene. Di conseguenza, per comprimere pastiglie cave alle alte densità necessarie, bastano pressioni minori e quindi minori intensità laser. (Dovrebbe essere possibile ottenere gusci liquidi cavi di DT con un processo analogo a quello di formazione delle bolle).

Se si vuole comprimere il combustibile da fusione di un fattore 10 000, quale grado di simmetria sferica deve avere l'implosione? Nel processo il raggio della pastiglia viene ridotto di un fattore 20. Qualsiasi variazione locale della velocità di implosione produrrà un errore proporzionale nella distanza percorsa dal materiale compresso; così un errore di una parte su 20 nella velocità produrrà un errore grande come il raggio della pastiglia compressa, con il risultato che la pastiglia compressa non sarà sferica. Per poter ottenere un alto grado di simmetria sferica durante una riduzione del raggio di un fattore 20, gli errori temporali e spaziali nella pressione di implosione applicata non devono superare l'1%.

La simmetria necessaria per l'implosione viene ottenuta irradiando la pastiglia uniformemente e simultaneamente con fasci laser multipli. L'atmosfera che si crea immediatamente attorno alla pastiglia irradiata completa l'opera: la diffusione degli elettroni nell'atmosfera fa da filtro alle residue disuniformità nell'ingresso laser. Complessivamente l'effetto dell'atmosfera sulla simmetria dell'implosione può essere paragonato a una Terra circondata da un'atmosfera nuvolosa di 6000 chilometri di spessore, riscaldata da 12 Soli disposti simmetricamente, ciascuno con un diametro apparente molte volte maggiore di quello del nostro vero Sole. In queste condizioni l'illuminamento dovrebbe essere così uniforme da eliminare praticamente le ombre. I calcoli bidimensionali eseguiti con elaboratori mostrano che l'atmosfera attorno alla pastiglia può ridurre gli errori di irraggiamento fino a un fattore 1000. Di conseguenza dovrebbe essere possibile ottenere una sfericità sufficiente per l'ignizione e la propagazione di un fronte di combustione da fusione in una pastiglia compressa.

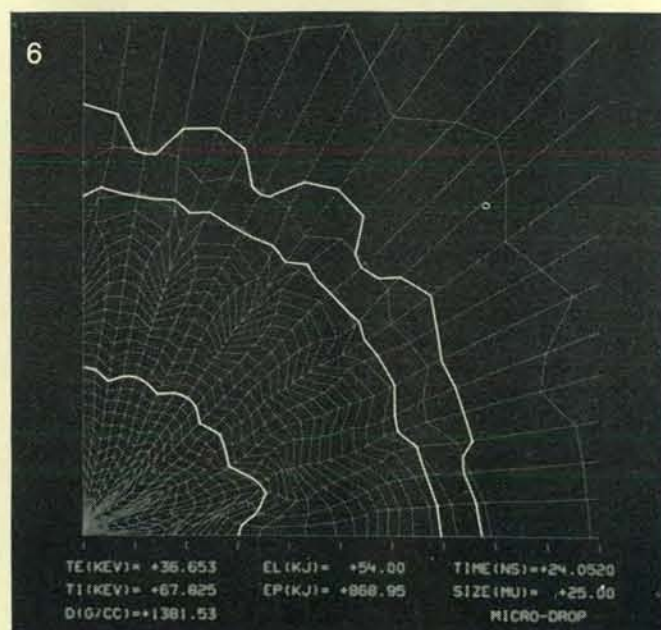
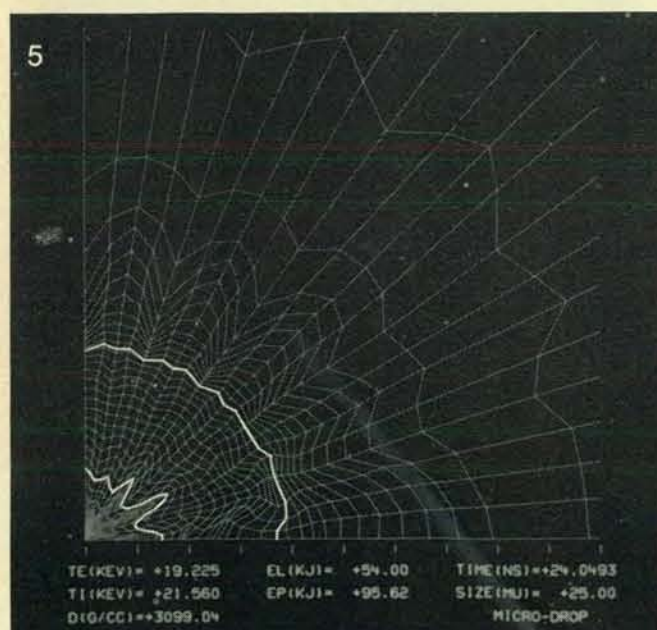
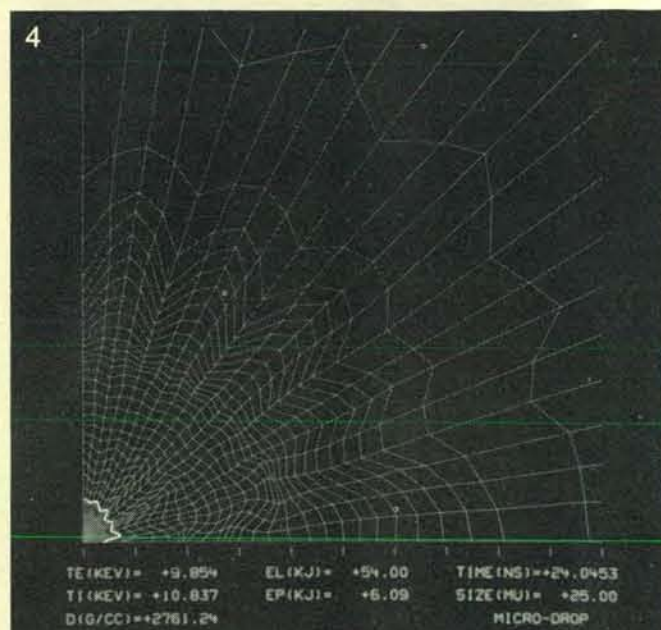
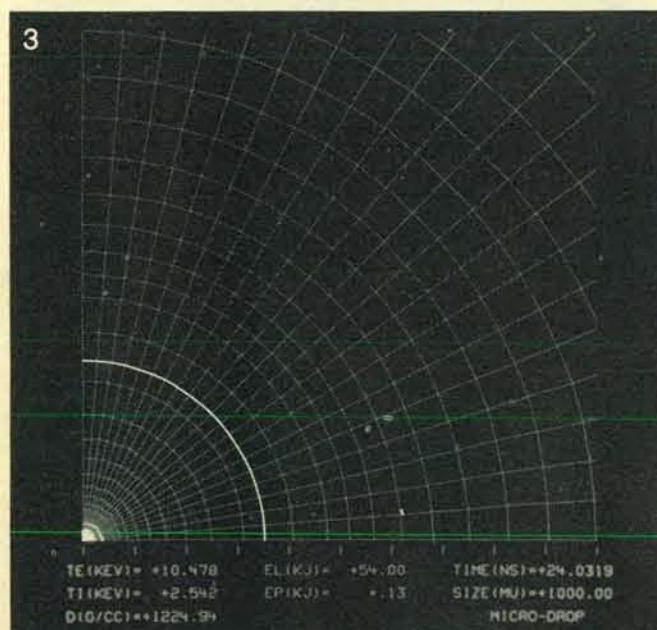
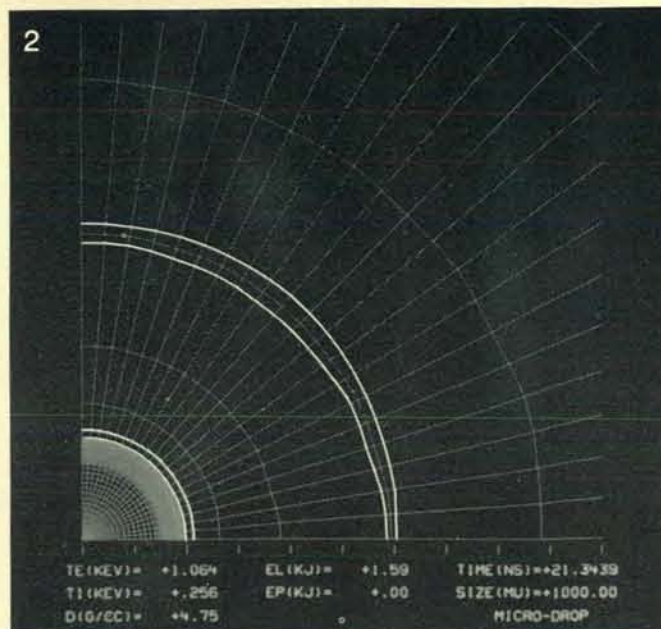
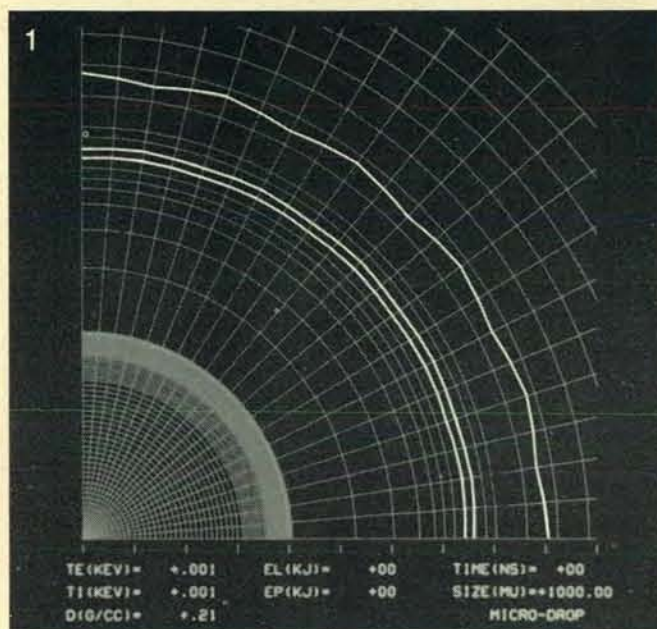
L'energia laser necessaria per un reattore a fusione viene fortemente ridotta dalla propagazione della reazione di fusione ottenibile in una pastiglia fortemente compressa. Le curve mostrano un'influenza della compressione della pastiglia sul rapporto tra la energia termonucleare liberata e l'energia laser impiegata per laser di diverse energie. Le curve colorate comprendono la propagazione, mentre quelle nere descrivono l'ignizione uniforme (senza propagazione). Ad altissime compressioni il guadagno diminuisce poiché l'energia necessaria per la compressione diventa eccessiva. Negli altri casi il guadagno aumenta con la compressione poiché il combustibile DT brucia con rendimento crescente e basta l'ignizione di una porzione più piccola di pastiglia per mantenere la propagazione radiale. Per una produzione di potenza un laser col 10 % di rendimento deve fornire circa 3×10^5 joule per ottenere un guadagno accettabile di 75:1. Dei 3×10^5 joule, circa il 95 % viene consumato per l'ablazione della pastiglia di combustibile fino a qualche decimo di milligrammo. Restano circa $1,5 \times 10^4$ joule (dell'ordine di 10^7 joule per grammo) nella pastiglia, dopo aver fornito l'energia necessaria per la compressione e per innescare l'ignizione termonucleare. Le diverse curve sono calcolate per l'optimum di peso della pastiglia e di sagoma dell'impulso.

Sono stati messi a punto programmi per elaboratori di estrema complessità per calcolare le implosioni generate dal laser e le microesplosioni termonucleari. Tali programmi vengono impiegati per la progettazione delle pastiglie da fusione, per fornire orientamenti sulla progettazione dei laser e per mettere a punto calcoli sugli esperimenti laser-plasma. Il più grande e complesso di tali programmi, denominato LASNEX, è stato messo a punto presso il laboratorio Lawrence di Livermore e comprende il calcolo del trasporto e delle interazioni dei fotoni laser, degli elettroni, degli ioni, dei raggi X e dei prodotti della reazione di fusione, oltre ai campi magnetici e elettrici e al comportamento idrodinamico della pastiglia (*si vedano le figure alle pagine 19 e 24*).

Per una data energia laser, il guadagno (il rapporto tra l'energia di fusione liberata e l'energia all'ingresso) è determinato da quattro fattori: il rendimento di combustione, l'energia di ignizione (corretta per la propagazione), l'energia di compressione e il rendimento di implosione. Il nostro modello matematico di un reattore a fusione laser prevede un guadagno di 75 se si può disporre di un rendimento laser del 10%. Per raggiungere tale guadagno, l'energia laser all'ingresso deve essere di 300 000 joule in un impulso accuratamente modellato dalla durata di circa un nanosecondo. Con tale energia all'ingresso il guadagno raggiunge il massimo quando la compressione diventa pari a 10 000 volte la densità allo stato liquido (*si veda la figura nella pagina a fronte*). Dato un laser da 300 000 joule che funziona al 10% di rendimento, si può sperare di riuscire alla fine a realizzare un impianto di potenza da 1000 megawatt (un gigawatt) nel quale gli impulsi laser danno luogo a 100 microesplosioni di fusione al secondo.

Oggi non esistono ancora laser dotati di tali prestazioni. Riteniamo, tuttavia, che la tecnologia attuale sia in grado di dimostrare la realizzabilità scientifica della fusione laser, ottenendo guadagni maggiori o uguali a 1. I calcoli indicano la necessità di energie laser dell'ordine di 5000 joule. Almeno quattro differenti tipi di laser sono oggi in grado di produrre impulsi

La compressione del combustibile nella fusione laser richiede l'applicazione di pressioni superiori a quelle esistenti al centro del Sole: circa 10^{12} (mille miliardi) di atmosfere. A una tale pressione gli isotopi liquidi dell'idrogeno possono essere compressi di un fattore 10^4 (*curva a sinistra*). Nella compressione di una pastiglia di combustibile a tale densità l'energia radiante del laser viene convertita in energia cinetica quando la materia viene accelerata ad alte velocità, quindi in energia interna quando la materia viene bruscamente arrestata. La minima velocità di implosione necessaria per raggiungere una compressione di 10^4 è di circa 3×10^7 centimetri al secondo, pari a un millesimo della velocità della luce (*curva a sinistra*). Il prodotto della pressione per la velocità fornisce l'intensità di ogni compressione. Per una compressione di 10^4 è necessaria un'intensità di 10^{19} watt per centimetro quadrato (*curva a destra*).

Un aumento crescente dell'intensità di potenza fino a 10^{19} watt per centimetro quadrato può essere ottenuto in quattro fasi. L'amplificatore laser (*a*) concentra energia sia nello spazio sia nel tempo da 10^5 a 10^{10} watt per centimetro quadrato. La focalizzazione del fascio laser (*b*) fornisce un ulteriore fattore da 10^4 a 10^5. L'atmosfera circostante la pastiglia bersaglio funziona da lente termica (*c*) trasportando l'energia all'ingresso alla minore superficie della pastiglia in ablazione e aumentando quindi l'intensità a 10^{15} — 10^{16} watt per centimetro quadrato. L'implosione (*d*) agisce sia come lente idrodinamica che come interruttore. Il movimento del materiale concentra l'energia cinetica su una superficie in continua diminuzione. Infine l'energia cinetica viene convertita in energia interna in un tempo molto minore di quello necessario per la produzione di energia cinetica da parte delle pressioni di implosioni applicate. Si raggiunge così, attraverso quattro fasi, l'intensità di 10^{19} watt per centimetro quadrato necessaria per la compressione.

superiori a un joule con durata uguale o minore di un nanosecondo; in ordine di sviluppo storico i loro elementi attivi sono il rubino, il vetro al neodimio, l'anidride carbonica e lo iodio. I primi due sistemi sono a stato solido, gli altri due gassosi.

Il laser a rubino deve essere scartato per la fusione, poiché il suo rendimento complessivo quando funziona a brevi impulsi arriva appena allo 0,001%. Inoltre è praticamente impossibile, oltre che estremamente costoso, procurarsi grossi pezzi di rubino sintetico aventi eccellenti proprietà ottiche.

Il vetro drogato con neodimio e «pompato» con lampada flash allo xeno emette una radiazione infrarossa con una lunghezza d'onda di 1,06 micromètri. I rendimenti nel funzionamento a brevi impulsi variano tra lo 0,02 e lo 0,1%, il che pone il laser a vetro al neodimio nel campo delle possibili applicazioni per esperimenti di laboratorio. Grossi pezzi di vetro al neodimio di eccellenti qualità ottiche sono costosi, ma facilmente reperibili. Uno degli aspetti interessanti di questo laser è che la sua lunghezza d'onda in uscita può essere dimezzata a 0,53 micromètri mediante l'effetto chiamato produzione della seconda armonica. Alle alte energie i ricercatori del centro francese di Limeil hanno ottenuto rendimenti maggiori del 50% nella conversione di lunghezza d'onda da 1,06 a 0,53 micromètri.

Dovrebbe essere possibile anche la produzione della quarta armonica, che dà origine a una radiazione ultravioletta di lunghezza d'onda 0,265 micromètri (2650 angstrom) con un rendimento complessivo che si avvicina al 25%. In tal modo si può usare un solo sistema laser per analizzare gli effetti delle diverse lunghezze d'onda sul processo di riscaldamento del plasma. In vari laboratori del mondo con sistemi laser al vetro al neodimio si sono ottenuti impulsi di energia tra 100 e 1000 joule con durate comprese tra 0,1 e 2 nanosecondi. Durante gli impulsi più brevi la potenza massima di uscita non focalizzata raggiunge da 0,1 a 3 terawatt (da 10^{11} a 3×10^{12} watt).

Negli ultimi due anni è notevolmente progredita la tecnologia dei laser ad anidride carbonica. Tali laser funzionano nell'infrarosso a 10,6 micromètri. Sono stati prodotti impulsi da 100 a 200 joule con durate da 1 a 2 nanosecondi (a un livello di potenza di picco di 100 gigawatt). Sono stati ottenuti rendimenti per impulsi brevi tra l'1 e il 2% e sembrano raggiungibili valori tra il 5 e il 10%. Non sono stati ancora ottenuti impulsi di durata inferiore a un nanosecondo. Lo sforzo maggiore nel programma di fusione laser dell'AEC con i laser ad anidride carbonica è quello prodotto attualmente dai nostri colleghi del laboratorio scientifico di Los Alamos, che stanno mettendo a punto un laser ad anidride carbonica da 10 000 joule e impulsi della durata di un nanosecondo.

Il laser a iodio funziona a una lunghezza d'onda di 1,34 micromètri. Lo iodio atomico viene preparato in un opportuno stato eccitato dalla fotodissociazione di composti gassosi quali lo iodotrifluoremetano (CF_3I) con lampade flash allo xeno. Tale sistema è oggetto di un profondo studio all'istituto Max Planck di fisica dei plasmi di Garching, nella Germania Occidentale. Come nel caso dell'anidride carbonica, il mezzo laser è poco costoso e quindi molto conveniente. I laser a iodio hanno mostrato un rendimento di circa lo 0,5%. Essi sono però di difficile controllo, poiché lo iodio ha un alto coefficiente di guadagno, il che significa che gli atomi, una volta pompati a un livello eccitato, emettono facilmente in modo stimolato la loro radiazione caratteristica cadendo in uno stato energetico inferiore. Poiché i sistemi laser a impulsi brevi devono immagazzinare grandi quantità di energia prima di poter dar luogo all'amplificazione dell'impulso, alti coefficienti di guadagno in amplificatori di grande apertura presentano due tipi di difficoltà. La prima è chiamata superfluorescenza, e non è altro che la normale fluorescenza emessa spontaneamente dal materiale laser eccitato, amplificata dal guadagno del materiale stesso. Se il coefficiente di guadagno specifico è alto, la potenza dissipata per superfluorescenza potrebbe superare la massima intensità di pompaggio ottenibile. L'energia non può quindi venire efficacemente immagazzinata nell'amplificatore laser.

La seconda, e più grave, difficoltà è dovuta alla retroazione della radiazione di superfluorescenza rinviata all'amplificazione laser. La retroazione genera oscillazioni laser parassite all'interno dell'amplificatore laser. La rapidità di emissione stimolata da tali oscillazioni parassite potrebbe superare di gran lunga qualsiasi intensità di pompaggio e limitare perciò la massima energia immagazzinata ottenibile. Sebbene possa sembrare quasi paradossale, per grandi sistemi di amplificatori laser sono necessarie transizioni laser con bassi coefficienti di guadagno.

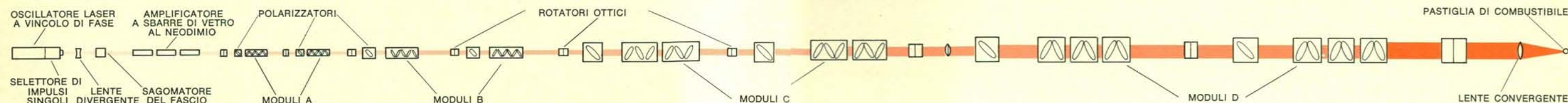
L'esplosione di una pastiglia di DT simulata da un programma di calcolo elaborato a Livermore e chiamato LASNEX. Sei fotogrammi di un film ricavati dall'elaboratore per un calcolo standard sono riprodotti nella pagina a lato. (L'ultimo fotogramma rappresenta un quadrante dello stesso stadio dell'esplosione rappresentato completo a pag. 19). I numeri sotto ad ogni fotogramma indicano in kiloelettronvolt (keV) la massima temperatura degli elettroni (*TE*) e degli ioni (*TI*). Un keV corrisponde a circa 10^7 kelvin. *D* è la densità massima in grammi per centimetro cubo. *EL* è l'energia laser complessiva all'ingresso, mentre *EP* è l'energia di fusione all'uscita, entrambe in kilojoule. Il tempo è espresso in nanosecondi e la larghezza del fotogramma in micromètri. (Si noti un fattore di cambiamento di scala 40 tra il fotogramma 3 e il 4). Le linee bianche nei primi tre fotogrammi sono isoterme della temperatura degli elettroni, mentre quelle negli altri tre fotogrammi sono isoterme della temperatura degli ioni. Nel fotogramma 1, ripreso all'istante zero, una pastiglia da 400 micromètri alla normale densità liquida è circondata da un'atmosfera a bassa densità che si estende oltre i 1000 micromètri. La luce laser viene assorbita tra le due isoterme molto vicine tra loro a un raggio di 750 micromètri e a una temperatura di 3×10^6 kelvin. L'isoterma a 900 micromètri è a 10^6 kelvin. Gli elettroni caldi non sono ancora penetrati nella superficie della pastiglia. Nel fotogramma 2 la pastiglia è stata compressa a 200 micromètri e la luce laser viene assorbita tra due isoterme a 10^7 kelvin a 600 micromètri. Nel fotogramma 3 è stata liberata tutta l'energia di 54 kilojoule del laser e la pastiglia è stata compressa a un raggio inferiore a 15 micromètri. Nel fotogramma 4 è indicato l'innesco dell'ignizione termonucleare con l'isoterma di temperatura degli ioni a 10^8 kelvin nel centro. Nel fotogramma 5, scattato quattro picosecondi dopo, il fronte di combustione a 10^8 kelvin si è propagato attraverso quasi tutta la pastiglia e un'isoterma raggiata a temperatura doppia ha cominciato a spostarsi verso l'esterno dal centro. Nel fotogramma 6 un'isoterma a 5×10^8 kelvin si è spostata di 7,5 micromètri dal centro ed è stata liberata metà dell'energia di fusione nucleare. La restante energia verrà prodotta nei successivi 12 picosecondi (0,012 nanosecondi). Un calcolo di questo tipo richiede parecchie ore di lavoro di quello che è oggi il più potente elaboratore del mondo, il CDC 7600. Vengono calcolate sessanta variabili a 2000 punti nello spazio e 10 000 punti nel tempo per fornire un totale complessivo di più di un miliardo di numeri.

Fino al momento attuale i sistemi laser a vetro al neodimio sono stati preferiti per dimostrare la realizzabilità scientifica della fusione laser in tutti i maggiori laboratori del mondo dove si sta studiando tale processo, tranne uno. Al laboratorio Lawrence di Livermore stiamo mettendo a punto diversi sistemi laser a vetro al neodimio per esperimenti di fusione di breve durata. I sistemi, tutti simili tra loro, sono costituiti da un oscillatore laser a bassa energia e da una cascata di amplificatori laser (*si vedano le figure nelle due pagine seguenti*).

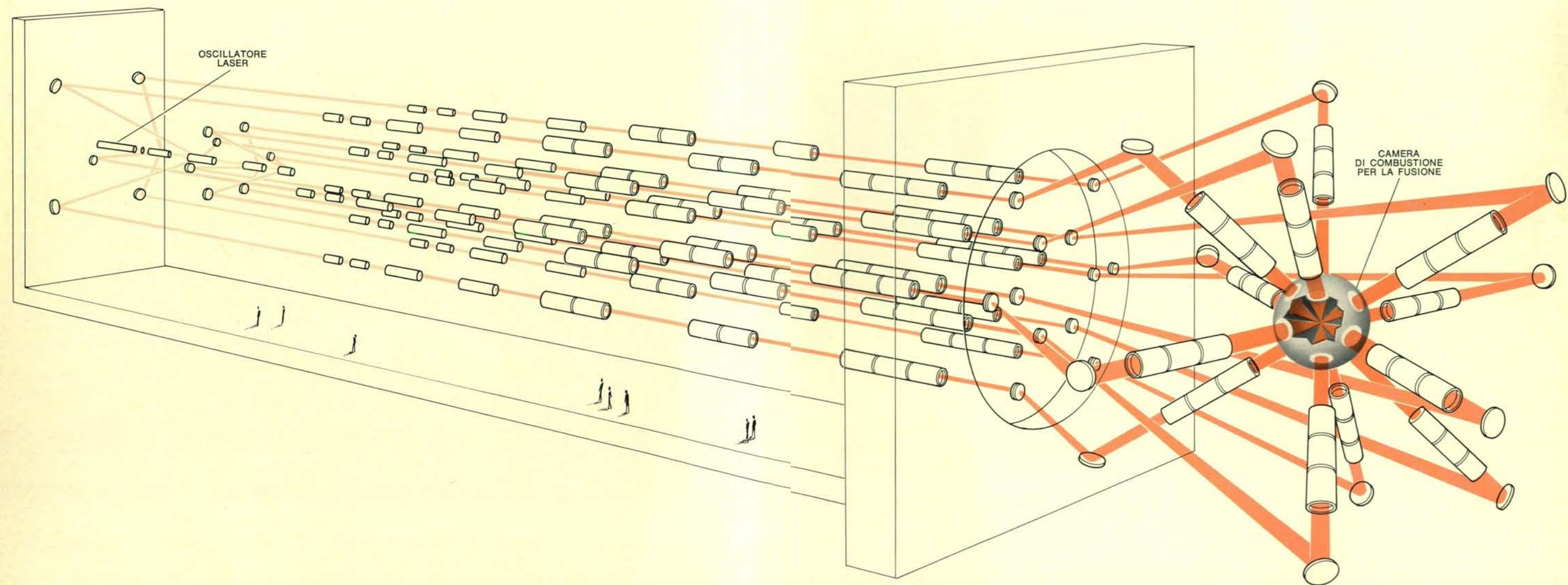
L'oscillatore produce un treno comprendente da 5 a 30 impulsi la cui durata è regolabile nell'intervallo tra 20 e

1000 picosecondi (millesimi di miliardesimo di secondo). Per produrre un impulso singolo soggetto ad amplificazione viene usato un otturatore elettroottico rapido. Prima di essere amplificato, l'impulso, che contiene solo circa 10^{-3} joule di energia laser, viene modellato nello spazio (e anche nel tempo, se si vuole). Poiché si richiedono energie totali di fascio singolo da 100 a 1000 joule, la cascata degli amplificatori deve fornire un guadagno ottico variabile da 100 000 a un milione.

Il massimo guadagno in ogni stadio della cascata è determinato da effetti ottici non lineari, i quali limitano la massima intensità della radiazione laser che può propagarsi nella materia. La non linearità nasce poiché l'intenso campo elettrico dell'onda luminosa produce un piccolo ma significativo aumento dell'indice di rifrazione del materiale ottico. La distruzione della coerenza spaziale e l'autofocalizzazione del fascio laser prodotte da tali effetti non lineari limitano la densità di potenza ottica del vetro tra 5 e 20 gigawatt per centimetro quadrato. Se il sistema deve mantenersi al disotto di questo limite, si deve aumentare l'apertura (cioè la sezione del fascio) degli amplificatori tra uno stadio e quello successivo. Un primo fattore 1000 del guadagno può essere ottenuto con amplificatori a sbarrette di vetro al neodimio con diametro tra i due e i quattro centimetri. Altri fattori da 100 a 1000 nel guadagno si ottengono con amplificatori di diametro oscillante tra i 10 e i 30 centimetri. Le unità di grande superficie vengono solitamente ricavate da una serie di dischi di vetro al neodimio anziché da una sola sbarra spessa. La perdita di luce dovuta alla riflessione superficiale dai dischi viene eliminata inclinandoli a un angolo chiamato angolo di Brewster e polarizzando opportunamente il fascio laser. Tale disposizione consente di pompare uniformemente i grandi amplificatori attraverso grandi aperture. I dischi sono completamente circondati da lampade flash allo xeno, che inviano la loro energia alle facce inclinate dei dischi.

A Livermore abbiamo collaudato con successo amplificatori con aperture fino a 20 centimetri. Un sistema a due fasci della potenza di un terawatt (10^{12} watt) viene normalmente impiegato

Una catena del sistema laser di Livermore sarà lunga 50 metri e sarà costituita da un piccolo laser, o oscillatore, per la creazione dell'impulso iniziale cui seguono 11 stadi di amplificazione. I primi tre stadi sono sbarrette piene di vetro al neodimio, mentre gli stadi successivi sono costituiti da dischi di vetro al neodimio di diametro crescente. Gli amplificatori a disco indicati con *A*, *B*, *C* e *D* hanno rispettivamente aperture di 5, 9, 20 e 30 centimetri. Ogni amplificatore a disco comprende un polarizzatore e un rotatore ottico di Faraday per far ruotare di 45° il piano di polarizzazione. Il sistema polarizzatore-rotatore garantisce che la luce riflessa del bersaglio DT non ritorni indietro nel sistema di amplificazione a elevato guadagno.

OSCILLATORE LASER

CAMERA DI COMBUSTIONE PER LA FUSIONE

Un impianto laser multifascio di cui è previsto il completamento nel 1977 presso il laboratorio Lawrence di Livermore, è stato progettato per dimostrare la possibilità di innescare microesplosioni termonucleari mediante l'implosione di pastiglie di isotopi dell'idrogeno. Tale impianto comprenderà 12 catene di amplificatori come quella della figura in alto o, a secondo del costo, 20 catene leggermente meno potenti. L'azione combinata dei fasci potrà irradiare pastiglie campione con 10 kilojoule di energia ottica in tempi da 100 a 500 picosecondi, equivalenti a una potenza di picco in uscita da 20 a 100 terawatt. La potenza, l'energia e la sagoma dell'impulso laser dovranno essere molto flessibili per poter provare diversi tipi di pastiglie.

nei primi studi dell'interazione laser-bersaglio. L'apertura finale di questo sistema è di nove centimetri. Entro il 1975 si prevede che debba entrare in funzione un sistema tra i 5 e i 10 terawatt comprendente da 6 a 12 fasci laser disposti a simmetria sferica. Un siffatto sistema dovrebbe produrre implosioni del mezzo ad alta compressione e facilitare la messa a punto di tecniche diagnostiche di implosione.

Il passo successivo dovrebbe essere il progetto e la realizzazione di un impianto a simmetria sferica i cui fasci combinati, prima di essere focalizzati trasportano da 50 a 100 terawatt di energia radiante. Quando i fasci vengono focalizzati, essi sono in grado di fornire 10 000 joule di energia a una pastiglia di combustibile di fusione in impulsi di durata compresa tra 100 e 500 picosecondi. Con alcuni ritocchi dovrebbe essere possibile fornire 50 000 joule in un impulso di pochi nanosecondi. Un siffatto impianto dovrebbe costare circa 20 milioni di dollari.

Un grande laser è costituito da sistemi comprendenti da 12 a 20 amplificatori in parallelo comandati da un solo oscillatore per garantire una sincronizzazione pressoché perfetta (*si veda la figura alle pagine 26 e 27 in basso*). La scelta di 12 o 20 fasci per ottenere i 10 000 joule di energia sul bersaglio è imposta da considerazioni di simmetria dell'implosione e dal costo di rifinitura dei grandi elementi ottici. Il progetto è attualmente in fase di realizzazione e una delle catene di amplificatori è quasi a punto. Con questo impianto, il cui completamento è previsto per l'inizio del 1977, riteniamo debba essere possibile dimostrare la realizzabilità scientifica della fusione laser raggiungendo o addirittura superando le condizioni in cui la resa di energia termonucleare eguagli l'energia ottica fornita all'ingresso.

Anche se un sistema laser con vetro al neodimio dovrebbe riuscire a dimostrare la possibilità della fusione laser, sappiamo che esso non può essere impiegato per un reattore a fusione. Un impianto di potenza della capacità di un gigawatt elettrico che bruci combustibile deuterio-trizio richiede un laser di circa 300 000 joule funzionante con una frequenza di ripetizione di 100 impulsi al secondo (o parecchi laser funzionanti a una frequenza di ripetizione inferiore) con un rendimento del 10%. Se il combustibile è sotto forma di pastiglie solide il laser deve fornire i suoi 300 000 joule a un livello di potenza di picco di 1000 terawatt; se il combustibile è sotto forma di pastiglie liquide cave, il livello di potenza può

GENERAZIONE DI IMPIANTI DI POTENZA TERMONUCLEARI	COMBUSTIBILI DA FUSIONE	PRODOTTI DI FUSIONE
PRIMA	DEUTERIO + TRIZIO ⟶	ELIO 4 + NEUTRONE
SECONDA	DEUTERIO + DEUTERIO <	ELIO 3 + NEUTRONE TRIZIO + IDROGENO
TERZA	BORO 11 + IDROGENO ⟶	ELIO 4 (TRE ATOMI)

I combustibili per impianti di potenza di fusione possono diventare meno costosi e a combustione più pulita se lo sviluppo della tecnologia laser permetterà di raggiungere densità ancora superiori a quelle necessarie per la fusione del deuterio e trizio. Poiché i combustibili per impianti più progrediti bruciano meno rapidamente, per una efficace combustione è necessario che il prodotto del raggio della pastiglia per la densità superi di un fattore da 10 a 100 quello richiesto per il combustibile deuterio-trizio.

ridursi fino a meno di 100 terawatt.

Se il sistema laser ha un rendimento del 10%, nel mezzo laser si hanno dispersioni di calore con una potenza media di 300 megawatt. Il mezzo laser, allo scopo di consentire un rapido flusso termico per l'asportazione del calore disperso, anziché una lenta conduzione termica, deve perciò essere un fluido. La grandezza del coefficiente ottico non lineare caratteristico della maggior parte dei liquidi suggerisce l'impiego di un mezzo laser gassoso dato che il laser deve funzionare ad altissime densità di potenza ottiche. D'altra parte i gas sono soggetti a scariche luminose. Una semplice teoria del processo prevede che la soglia per la scarica luminosa è inversamente proporzionale al quadrato della lunghezza d'onda: le lunghezze d'onda inferiori sono dunque da preferirsi a quelle maggiori. Per una lunghezza d'onda di 10,6 micromètri la soglia di scarica misurata a pressione di circa un'atmosfera è dell'ordine di 10^9 watt per centimetro quadrato; per una lunghezza d'onda di 1,06 micromètri la soglia aumenta di un fattore 100. Pertanto la scarica luminosa nel mezzo laser e le instabilità del plasma nel bersaglio suggeriscono l'impiego di radiazione di lunghezza d'onda minore. Però, se la lunghezza d'onda è talmente piccola da consentire il simultaneo assorbimento di due fotoni da parte di un atomo o di una molecola del mezzo laser, con conseguente fotoionizzazione, il fascio laser, anziché essere amplificato, verrà fortemente assorbito. La miglior lunghezza d'onda per un impianto di potenza a fusione laser è compresa tra 0,3 o 0,8 micromètri circa.

Se la pressione del mezzo gassoso nel nostro ipotetico sistema laser è inferiore a qualche atmosfera, il sistema dovrebbe poter raggiungere un livello di potenza fino a 5×10^{10} watt per centimetro quadrato: di conseguenza, una potenza totale di 10^{14} watt (per bersaglio a pastiglia cava) o di 10^{51} watt (per bersaglio solido) richiede un sistema laser con un'apertura finale di 2000 o di 20 000 centimetri quadrati rispettivamente. Se si impiegano 12 fasci laser per ottenere la necessaria simmetria di irraggiamento, basta che i singoli amplificatori finali abbiano diametri compresi tra i 15 e i 45 centimetri, dimensioni facilmente raggiungibili allo stato attuale della tecnologia ottica.

Per concludere, un sistema laser da utilizzarsi per un impianto di potenza di fusione deve soddisfare i seguenti requisiti: la lunghezza d'onda di funzionamento deve essere compresa tra 300 e 800 nanometri, il mezzo deve essere un gas a pressione inferiore a qualche atmosfera e, per limitare la superfluorescenza e le oscillazioni parassite, il gas deve avere un basso coefficiente di guadagno specifico. Dobbiamo perciò cercare una debole transizione elettronica in un atomo o in una molecola. Inoltre si deve poter raggiungere un rendimento all'incirca del 10%.

La ricerca di un nuovo mezzo laser è attualmente in fase di sviluppo nel nostro laboratorio, presso il laboratorio scientifico di Los Alamos, al laboratorio Sandia di Albuquerque – entrambi dell'AEC – e in molti altri laboratori del mondo. Un tipo di transizione elettronica in esame è quella dell'ossigeno atomico a 0,5577 micromètri, riga spettrale responsabile del colore verde dell'aurora boreale. Altri elementi che nella scala periodica si trovano nella stessa colonna dell'ossigeno (zolfo, selenio e tellurio) mostrano transizioni dello stesso tipo anch'esse attualmente all'esame. Siamo dell'avviso che, anche se il mezzo laser ideale per un impianto di potenza di fusione non è ancora stato trovato, la

DENSITÀ DEL COMBUSTIBILE × RAGGIO (GRAMMI/CM²)	DENSITÀ DEL COMBUSTIBILE (GRAMMI/CM³)	IMPULSO LASER NECESSARIO (JOULE/NANOSECONDI)	RENDIMENTO LASER (%)	ENERGIA PERCENTUALE LIBERATA IN: RAGGI X	PARTICELLE CARICHE	NEUTRONI
2 — 5	10^3 — 4 × 10^3	$10^5/3$	~ 10	2 — 10	25 — 30	65 — 75
10 — 20	~ 10^4	$10^6/1$	⩾ 10	20 — 50	30 — 60	20 — 45
~ 200	~ 3 × 10^5	$10^8/0,3$	⩾ 30	30 — 70	30 — 70	< 0,1

Il rendimento del sistema viene ottimizzato aumentando sia il raggio della pastiglia che la densità del combustibile per raggiungere i valori voluti. Ciò richiede a sua volta impulsi laser più brevi e di maggiore energia, prodotti con maggior rendimento laser. I combustibili proposti per impianti di potenza della seconda e della terza generazione producono quantità di energia sempre minori sotto forma di neutroni e questo è un fatto positivo, se si tiene conto che la radioattività prodotta sulle pareti della camera di combustione è all'incirca proporzionale all'emissione di neutroni. La reazione di un protone e di un isotopo del boro, il boro 11, pare offrire buone prospettive per la terza generazione degli impianti di fusione termonucleare.

sua ricerca e le relative tecnologie costituiscano un compito realizzabile, anche se faticoso.

Un impianto di potenza di fusione col laser, oltre a sostanziali progressi nella tecnologia laser, richiederà la soluzione di molti altri problemi tecnici. Le detonazioni ad alto rendimento delle pastiglie di combustibile da fusione per la produzione pratica di elettricità avviene in una scala dei tempi dell'ordine di 10^{-11} secondi o meno ancora. Poiché l'energia liberata dovrebbe essere di almeno 10^7 joule, la potenza di fusione prodotta dovrebbe raggiungere un picco di 10^{18} watt (10^7 diviso per 10^{-11}). Un tale ritmo di produzione di energia (che è certamente intermittente) è un milione di volte maggiore della potenza di tutti i macchinari mai realizzati dall'uomo ed è circa 10 volte maggiore della potenza radiante totale della luce solare incidente su tutta la Terra. La sfida tecnologica della fusione laser sta nel riuscire a costruire un impianto di potenza attorno a microesplosioni di fusione con picchi astronomici di potenza che possano ripetersi per molti anni da decine a centinaia di volte al secondo. Anche se la cosa sorprende, è possibile farlo.

L'energia liberata in una microesplosione deuterio-trizio viene trasportata da neutroni, raggi X e particelle cariche. Ciascuna di tali radiazioni rappresenta una particolare minaccia all'integrità del « primo schermo », o superficie più interna, della camera di combustione. Le singole particelle della radiazione escono con uno spettro energetico in un ampio intervallo di velocità, fino al valore della velocità della luce. La dispersione delle velocità è una circostanza fortunata perché impedisce alle particelle di colpire tutte contemporaneamente il primo schermo. Questo fatto riduce di molto la potenza di picco con cui viene ceduta energia e rende possibile una lunga sopravvivenza dello schermo.

Il pericolo maggiore per lo schermo è costituito dai neutroni, che trasportano circa i tre quarti dell'energia di fusione prodotta. Sebbene i neutroni ad alta energia attraversino la materia molto più rapidamente di qualsiasi altra radiazione, essi lasciano tuttavia tracce del loro passaggio sotto forma di dislocazioni e disintegrazioni di atomi. Le dislocazioni, che strappano gli atomi via dalla loro posizione nel reticolo cristallino, sono il risultato di collisioni simili a quelle delle palle da biliardo che si hanno quando i neutroni attraversano il primo schermo. In qualche caso i neutroni frantumano qualche nucleo atomico dello schermo espellendo un protone, un neutrone, un deutone (nucleo di deuterio), un trizione (nucleo di trizio) o un nucleo di elio. Tutte le particelle espulse (a eccezione dei neutroni secondari) formano atomi gassosi all'interno dell'interno della parete quando rallentano dopo la nascita. Gli atomi alla fine si raggruppano in piccolissime bolle gassose le cui pressioni possono salire fino a centinaia di atmosfere prima di provocare la rottura della superficie dello schermo. Inoltre, il nucleo frantumato è solitamente radioattivo e va ad arricchire il carico radioisotopico dell'impianto di potenza, che è costituito principalmente dai radioisotopi prodotti quando i neutroni primari vengono catturati alla fine dai nuclei atomici dello schermo.

I problemi di degradazione provocata dai neutroni pongono un limite superiore alla vita media del primo schermo in tutti gli impianti di potenza di fusione che bruciano DT, siano essi del tipo a laser o a confinamento magnetico. Gli esperimenti preliminari indicano che il miglior materiale disponibile per lo schermo può resistere al bombardamento con neutroni da 14 MeV (quelli liberati nella fusione deuterio-trizio) per qualche dozzina di anni, purché la potenza del flusso neutronico venga mantenuta al di sotto di un megawatt per metro quadrato. Ciò implica che qualsiasi reattore a fusione da un gigawatt che usi come combustibile DT richiede una superficie del primo schermo di 1000 metri quadrati, in una sola camera o in più camere. Nel caso del confinamento magnetico, se, per aumentare la superficie del primo schermo, si aumenta il volume della camera, si va incontro a un grave inconveniente, poiché i magneti devono essere all'esterno dello schermo; se li spostiamo all'esterno essi devono aumentare in proporzione di dimensione (e quindi divengono più costosi) per produrre lo stesso campo magnetico nel centro della camera.

Il pericolo costituito dai raggi X per la sopravvivenza del primo schermo è meno preoccupante. Poiché la maggior parte dei fotoni X hanno energie maggiori di 5000 elettronvolt, essi penetrano in profondità nella parete depositandovi la loro energia senza danno. Sebbene l'energia trasportata dai raggi X con meno di 2000 elettronvolt sia meno dello 0,1% dell'energia di una microesplosione, tali raggi X depositano la loro energia nello spessore di un micromètro della superficie dello schermo, portano rapidamente ad alta temperatura una sottile pellicola di essa. L'elevato gradiente termico e gli intensi sforzi meccanici prodotti in questo modo possono provocare ad ogni microesplosione un lieve sfaldamento dello schermo. Una soluzione possibile è l'impiego per la costruzione del primo schermo di materiali quali il litio, il berillio, il carbonio, che sono relativamente trasparenti ai raggi X. Questi

ultimi potrebbero infatti penetrare molto in profondità nello strato superficiale di tali materiali, riscaldando una massa maggiore a una temperatura inferiore e riducendo gli sforzi massimi a livelli accettabili.

Ben più complessi e sostanziali sono i pericoli portati dalle particelle cariche prodotte nelle microesplosioni all'integrità del primo schermo. Il primo è analogo a quello dei raggi X molli: gli ioni meno energetici (deutoni e trizioni termici), anch'essi assorbiti in uno strato sottile, possono provocare intensi sforzi termomeccanici in una sottile pellicola del primo schermo. Gli ioni più energetici sono invece abbastanza penetranti e scarsi di numero per cui presentano un pericolo relativamente modesto.

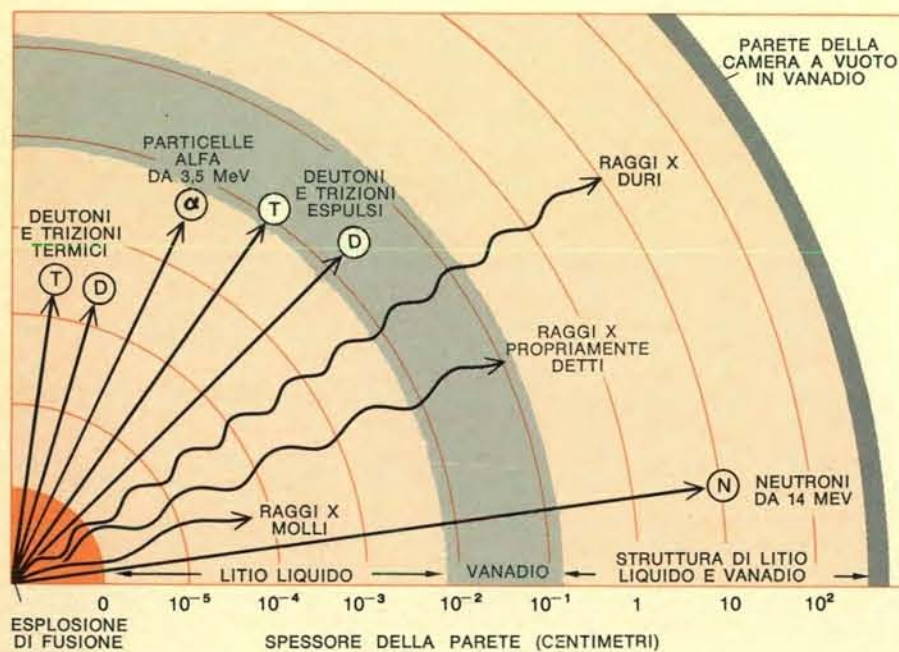
Il secondo più grosso pericolo rappresentato dalle particelle cariche presenta delle analogie con il problema delle bolle di gas associate ai neutroni: i nuclei energetici che vanno a morire nei primi micromètri della superficie catturano rapidamente elettroni e si trasformano in atomi gassosi che si concentrano in bolle di dimensioni crescenti con alte pressioni interne. Queste bolle possono frantumare la superficie di una qualsiasi parete solida in un tempo minore della durata di servizio desiderata per un impianto di potenza a fusione.

Un modo per ovviare agli assalti alla superficie più interna del primo schermo del reattore è il rivestimento della sua superficie con uno strato rigenerante di litio metallico liquido di spessore di qualche decimo di millimetro (*si veda la figura qui sotto*). Il litio potrebbe essere tenuto in posizione fissa dalla tensione superficiale originantesi su un guscio di supporto corrugato, per esempio di vanadio o di niobio. Una siffatta pellicola liquida dovrebbe essere esente da sforzi termomeccanici. Inoltre, gli atomi gassosi prodotti dall'azione dei neutroni e dalla neutralizzazione elettrica degli ioni di bassa energia dovrebbero migrare rapidamente verso la superficie interna della pellicola di litio liquido e potrebbero essere asportati insieme alla piccola quantità di litio vaporizzato durante ogni impulso di riscaldamento.

Contrariamente a quanto può suggerire l'intuizione, lo sforzo puramente meccanico applicato alla prima parete dai frammenti espulsi verso l'esterno in una microesplosione di fusione è estremamente piccolo. Per esempio, una microesplosione che libera circa 10 milioni di joule, paragonabile all'energia prodotta da due chilogrammi di un potente esplosivo chimico, applica alla camera di combustione termonucleare una pressione istantanea non superiore a quella che si ha in un petardo di dimensioni medie. La spiegazione di questo apparente paradosso sta nel fatto che la massa della pastiglia da fusione è solo di circa un milligrammo, ovvero inferiore di un fattore due milioni ai due chilogrammi di esplosivo chimico. Poiché l'impulso di un'onda esplosiva è proporzionale alla radice quadrata del prodotto tra l'energia e la massa associata all'onda, la piccolissima massa della pastiglia di confronto a quella dell'esplosivo chimico, implica una riduzione di un fattore 1400 nell'impulso di un'onda esplosiva di fusione rispetto all'impulso di un'onda d'urto provocata da un esplosivo chimico.

Il calore liberato negli impulsi dei raggi X e del plasma, che costituisce circa un quarto dell'energia totale di fusione di esplosioni DT, può essere asportato dal retro del primo schermo mediante una corrente refrigerante o la superficie evaporante di un radiatore. I neutroni, che trasportano circa tre quarti dell'energia di fusione, attraversano il primo schermo e vengono arrestati nello schermo successivo assorbente contenente litio; la reazione tra un neutrone e un atomo di litio 6 produce trizio ed elio; una parte del trizio viene rinviata al reparto di produzione di nuove pastiglie.

L'assenza di campi magnetici nel sistema laser consente di raffreddare lo schermo che produce trizio con fluidi, come il litio, che conducono elettricità; i fluidi conduttori presentano particolari problemi nei sistemi a confinamento magnetico.

Parecchie altre considerazioni risultano importanti nella progettazione dello schermo da disporre attorno a una camera di combustione per fusione. I fasci laser e le pastiglie di combustibile devono essere introdotti attraverso lo schermo lungo percorsi a zig-zag per impedire la fuga di raggi X e di neutroni (*si veda la figura a pagina 31*). Le pastiglie di combustibile potrebbero essere caricate elettricamente e guidate elettrostaticamente lungo un percorso a zig-zag, mentre i fasci laser possono essere guidati con specchi lungo analoghi percorsi a zig-zag. Lo specchio posto al termine di ogni percorso laser deve essere concavo per focalizzare la luce sulla pastiglia. È probabile che la superficie frontale dell'ultimo

Le radiazioni di fusione che colpiscono il primo schermo di una camera di combustione termonucleare penetrano a profondità variabili a seconda della loro energia. Lo schermo deve essere progettato in modo da rendere minimo il danno alla sua struttura. Circa il 70 % dell'energia liberata nella fusione del deuterio e del trizio viene trasportata da neutroni da 14 MeV (megaelettronvolt), circa il 25 % da particelle cariche (deutoni, trizioni e particelle alfa) e una piccola percentuale da raggi X. I deutoni e i trizioni che trasportano solo la normale energia di esplosione, corrispondente a 10^{9} kelvin, sono detti termici. I deutoni e i trizioni «espulsi» sono stati accelerati a energie circa 100 volte superiori attraverso urti con i neutroni nella zona dell'esplosione. La pellicola di litio liquido che ricopre la superficie del primo schermo arresta i raggi X più molli e tutte le particelle cariche capaci di generare bolle di gas radioattivo. Come materiale strutturale verrà probabilmente usato il vanadio dato che esso libera la minima radioattività sotto bombardamento neutronico rispetto a qualsiasi altro materiale resistente ad alte temperature che sia anche compatibile col litio liquido.

RIFORNIMENTO DI DEUTERIO
RIFORNIMENTO DI TRIZIO
LAVORAZIONE DELLE PASTIGLIE
CAMERA DI COMBUSTIONE
SISTEMA DI ASPORTAZIONE DEL CALORE
PRIMO SCHERMO
FASCI LASER CONVERGENTI
4 METRI
SCHERMO PER NEUTRONI
CONTENITORE A VUOTO PRIMARIO
ALL'IMPIANTO A VUOTO E ALLA SEPARAZIONE ISOTOPICA
FASCIO LASER
SPECCHIO
SEPARATORE DEI FASCI
AMPLIFICATORE LASER
PREAMPLIFICATORE LASER
OSCILLATORE LASER

Un impianto di potenza di fusione laser è concettualmente semplice. Pastiglie di deuterio e trizio liquidi di circa un millimetro di diametro vengono guidate elettrostaticamente su un percorso a zig-zag attraverso uno schermo per neutroni fino a cadere liberamente verso il centro della camera di combustione. Qui le pastiglie vengono irraggiate simmetricamente e compresse mediante fasci laser convergenti. I percorsi a zig-zag hanno la funzione di impedire la fuga di raggi X e di neutroni durante l'esplosione. Negli impianti di potenza di fusione della prima generazione il calore verrà rimosso con scambiatori termini tradizionali, quali un flusso di litio liquido, e utilizzato per produrre vapore e alimentare dei turbogeneratori. Negli impianti più progrediti potrà verificarsi una conversione diretta in elettricità di una grossa frazione dell'energia delle microesplosioni. Un impianto capace di innescare 100 microesplosioni al secondo potrebbe produrre tra 100 e 1000 megawatt di elettricità. Per una produzione economica di potenza il costo di ogni pastiglia non dovrebbe superare un cent (circa 7 lire).

specchio debba essere di metallo liquido continuamente sostituibile per potersi adattare all'ambiente nucleare. Lo specchio, però, non richiede necessariamente un elevato rendimento ottico, ma è sufficiente che la maggior parte della luce laser focalizzata cada su una macchia di meno di due millimetri da una distanza di uno o due metri.

I primi reattori a fusione laser impiegheranno probabilmente il combustibile deuterio-trizio, il quale richiede laser più piccoli. Però, se non fosse necessario il trizio, si potrebbero evitare gli schermi di litio per la rigenerazione del trizio. Per esempio, con laser da 10^6 joule si potrebbe impiegare come combustibile deuterio puro (DD). I prodotti della fusione del deuterio sono nel 50% degli eventi elio 3 e un neutrone, nel restante 50% trizio e un protone (*si veda la tabella nelle pagine 28 e 29*). La maggior parte di questo trizio brucia mentre la pastiglia compressa si consuma, ma il restante trizio si accumula nei frammenti della microesplosione e può essere poi recuperato dall'impianto a vuoto della camera di combustione e inviato alla « fabbrica » di pastiglie per venire miscelato al deuterio.

Alla fine si raggiunge uno stato di equilibrio nel quale in ogni pastiglia di deuterio è presente una piccolissima quantità costante di trizio, che ne facilita l'ignizione.

Il fatto di non dover rigenerare il trizio nello schermo semplifica la realizzazione e il funzionamento dello schermo stesso. Si può pensare, per esempio, a uno schermo per neutroni costituito da blocchi di grafite attraversati da tubi di raffreddamento la cui unica funzione è quella di trasformare in calore, per esempio a 2000 kelvin, l'energia dei neutroni; tale calore potrebbe essere asportato da elio ad alta pressione. Tale sistema sarebbe l'analogo per la fusione dei promettenti reattori raffreddati a gas ad alta temperatura (*High Temperature Gas Cooled Reactor*) che hanno raggiunto oggi la soglia di un impiego commerciale per la produzione di energia della fissione. Maggiore è la temperatura del fluido usato per azionare i turbogeneratori, maggiore è il rendimento termico dell'impianto di potenza e minore la dispersione di calore (inquinamento termico) nell'ambiente circostante.

Poiché un reattore a fusione non dovrebbe contenere grandi quantità di materiali radioattivi volatili e tossici, esso potrebbe funzionare a temperature ancora maggiori di quelle dei futuri reattori a fissione raffreddati a elio.

I neutroni della fusione possono essere anche utilizzati per la conversione di isotopi non fissili dell'uranio o del torio in isotopi fissili, proprio come avviene nei reattori a fissione del tipo autofertilizzante (*breeder*). I neutroni liberati nelle microesplosioni del DT o del DD sono in grado di produrre sotto forma di isotopi fissili dell'uranio o del plutonio una quantità di energia da 20 a 50 volte maggiore di quella liberata nella reazione di fusione originale.

Tali reattori ibridi a fusione-fissione potrebbero fare la loro comparsa prima che diventino economicamente competitivi i sistemi a fusione pura.

Al di là della fusione del deuterio puro, quali altre reazioni nucleari potrebbero essere prese in esame? Esiste un'attraente possibilità. Quando il prodotto densità-raggio della pastiglia compressa diventa superiore a qualche centinaio (il che implica densità di pastiglie 100 000 volte superiori a quella allo stato liquido) i raggi X emessi dal combustibile che brucia non possono più sfuggire subito e vengono efficacemente intrappolati. Se si potranno raggiungere le elevate densità necessarie per produrre questo effetto, certi combustibili da fusione che non subiscono l'ignizione nel modo caratteristico potranno ugualmente essere bruciati con adeguato rendimento. Per queste applicazioni sono però necessari laser con energie in uscita di 10^8 joule e rendimenti del 50%. Forse la reazione più interessante è quella tra l'idrogeno comune e il boro comune, il boro 11. Tale reazione potrebbe essere propriamente chiamata una reazione termonucleare di fissione poiché in essa vengono prodotte più particelle di quante ne vengono consumate. Il suo grande pregio è quello di convertire reagenti non radioattivi in prodotti non radioattivi. La reazione tra un protone e un nucleo di boro 11 avviene in tre fasi: nella prima il protone si unisce al nucleo di boro 11, formando carbonio 12 in uno stato eccitato. Il carbonio eccitato subisce immediatamente la fissione, trasformandosi in elio 4 e berillio 8, il quale, a sua volta, si fissiona istantaneamente in altri due nuclei di elio. I prodotti finali della reazione sono quindi tre atomi di elio di elevata energia. Reazioni collaterali, in effetti molto rare, nello 0,1% dei casi producono un neutrone a bassa energia e un nucleo debolmente radioattivo (carbonio 14).

Anche in tal caso un impianto di potenza basato sulla reazione di boro 11 e di un protone dovrebbe produrre una quantità di scorie radioattive 1000 volte minore di quelli che utilizzano la reazione DT.

Finora abbiamo descritto solo metodi convenzionali di trasporto del calore per produrre elettricità in un impianto di potenza a fusione laser. Non si potrebbe catturare in modo più diretto l'energia liberata nelle microesplosioni? Dopo tutto, disponiamo in partenza di cariche elettriche che schizzano via ad alta velocità dal punto dell'esplosione. Vorremmo arrivare a un flusso di cariche elettriche attraverso linee di trasmissione in reti elettriche ad alta tensione.

È davvero necessario trasformare e degradare l'energia lungo il percorso facendole attraversare una caldaia?

Forse no. Nelle microesplosioni con combustibile di deuterio la maggior parte dell'energia di fusione viene prodotta sotto forma di particelle cariche e la maggior parte dell'energia dei neutroni viene ceduta al plasma fortemente compresso. È ben noto che se un plasma viene fatto espandere in un campo magnetico, esso finisce col comprimere il campo e col respingerlo verso l'esterno. La compressione del campo magnetico può essere sfruttata per una conversione diretta in elettricità mediante il principio dell'induzione elettromagnetica scoperto da Faraday. Basta disporre le cose in modo che le linee in movimento del campo magnetico intersechino le spire di una bobina d'induzione. In esperimenti eseguiti con plasmi riscaldati da laser immersi in un campo magnetico si è visto che più del 70% di tutta l'energia fornita al plasma dal laser all'inizio viene trasferita al campo magnetico sotto forma di compressione delle linee del campo magnetico. Si è calcolato che in condizioni molto simili a quelle esistenti per le microesplosioni in deuterio si potrebbe assorbire nella compressione del campo magnetico dall'80% al 90% dell'energia iniziale di espansione nell'esplosione e che tale energia potrebbe essere poi estratta per induzione di Faraday dal campo magnetico per convertirla direttamente in elettricità.

Per 20 anni i tentativi di sfruttare i processi di fusione quali fonti di energia per l'umanità hanno seguito la strada del confinamento magnetico. Ora, per raggiungere lo stesso traguardo, è stato ideato un metodo del tutto indipendente, la fusione mediante implosione provocata dal laser. Nel cercare una soluzione di qualsiasi problema di grande importanza è sempre conveniente poter seguire più strade, poiché la diversità dei metodi aumenta la probabilità di raggiungere il successo.

L'equilibrio geochimico di crosta, atmosfera e oceani

Le zone esterne della Terra sono collegate mediante complessi cicli geochimici in un unico sistema nel quale i principali costituenti sono in equilibrio

di Raymond Siever

Per cercare di capire come funziona un sistema naturale vasto e complesso, quale è l'interazione della superficie solida della Terra con l'atmosfera e gli oceani, val la pena a volte di ignorare la massa dei particolari e di considerare in maniera nuova il sistema nel suo complesso. Questo tipo di prospettiva non consiste nel guardare la Terra dalla Luna (anche se il confronto con altri pianeti può essere utile), ma piuttosto nel stare qui sulla Terra e nel trarre fuori la natura dei processi generali, che funzionano su scala mondiale, dai meccanismi particolari che operano in determinati tempi e luoghi in componenti isolati del sistema. Forse il modo più efficace per costruire questo modello complessivo è di porre a confronto gli ingressi (*input*) e le uscite (*output*) dei principali componenti chimici del sistema considerato nel suo insieme e quindi dedurre i vari tipi di reazioni chimiche e di processi fisici che controllano l'equilibrio. Questa particolare forma di analisi è chiamata in gergo moderno « ingegneria dei sistemi generali »; più di un secolo prima però che l'ingegneria dei sistemi si chiamasse così, i geologi avevano adottato lo stesso metodo per trattare i problemi della Terra come per esempio l'origine delle glaciazioni.

In un'analisi di questo tipo, molto dipende da come vengono definiti i limiti del sistema e gli stati della materia nel sistema. Per lo scopo che mi propongo in questo articolo, gli elementi che includo sono l'atmosfera, gli oceani e tutte le altre masse d'acqua sulla superficie solida della Terra o nelle sue vicinanze, le rocce sedimentarie, metamorfiche e ignee della crosta terrestre, insieme alle parti del mantello superiore che interagiscono con la crosta (*si veda l'illustrazione in basso*). Le fasi fluide di questo sistema comprendono i gas dell'atmosfera, le quantità relativamente piccole di gas imprigionate nella crosta e l'acqua allo stato liquido contenuta negli oceani, nei laghi, nei fiumi di tutto il pianeta, insieme alla grande quantità di acqua che si trova nelle rocce sedimentarie in profondità. I componenti chimici del sistema di cui mi servirò come « tracce » comprendono molti degli elementi più abbondanti: sodio (Na), potassio (K), calcio (Ca), magnesio (Mg) e cloro (Cl), tutti generalmente presenti come ioni; carbonio (C), presente sia nel materiale organico, soprattutto nella formaldeide (CH_2O), o in carbonati, cioè l'anidride carbonica (CO_2) e l'ione carbonato (CO_3^{--}); il silicio (Si) soprattutto in forma di silice (SiO_2), l'idrogeno (H^+) e l'ossigeno molecolare (O_2). Trascurerò alcuni elementi importanti come lo zolfo, l'azoto e l'alluminio così come molti elementi meno abbondanti non perché non siano significativi, ma perché è possibile illustrare la natura generale del sistema con gli elementi che ho scelto.

I processi geofisici attraverso cui la crosta, l'atmosfera e l'oceano vengono in contatto tra di loro sono rappresentati schematicamente in questa sezione del guscio esterno della Terra. La crosta profonda e il mantello superiore sono connessi con l'atmosfera e l'oceano per mezzo dell'emanazione di gas e dell'estrusione di lava dai vulcani nonché dal sistema delle dorsali medio-oceaniche. Le rocce della crosta sono riciclate attraverso il mantello nelle zone di subduzione sottostanti le fosse oceaniche. I sedimenti sono riciclati quando sono sepolti in profondità e metamorfosati o fusi nella crosta profonda o nel mantello superiore. Le rocce ignee, metamorfiche e sedimentarie sono sollevate nelle catene montuose ed esposte all'azione dell'atmosfera e dell'idrosfera. I prodotti dell'erosione e dell'alterazione sono portati nell'oceano dalle acque sotterranee e superficiali.

In questo diagramma semplificato è rappresentato il sistema chiuso di reazioni che avvengono nella crosta terrestre, nell'atmosfera e negli oceani e che costituiscono un modello stazionario. I componenti principali del sistema sono due reattori chimici, uno caldo, profondo e non stagno, l'altro freddo e superficiale. I due reattori sono collegati da molte vie che sono controllate da differenze di temperatura e di pressione. Solo l'idrogeno e l'elio sono in grado di sfuggire al sistema a causa della loro leggerezza.

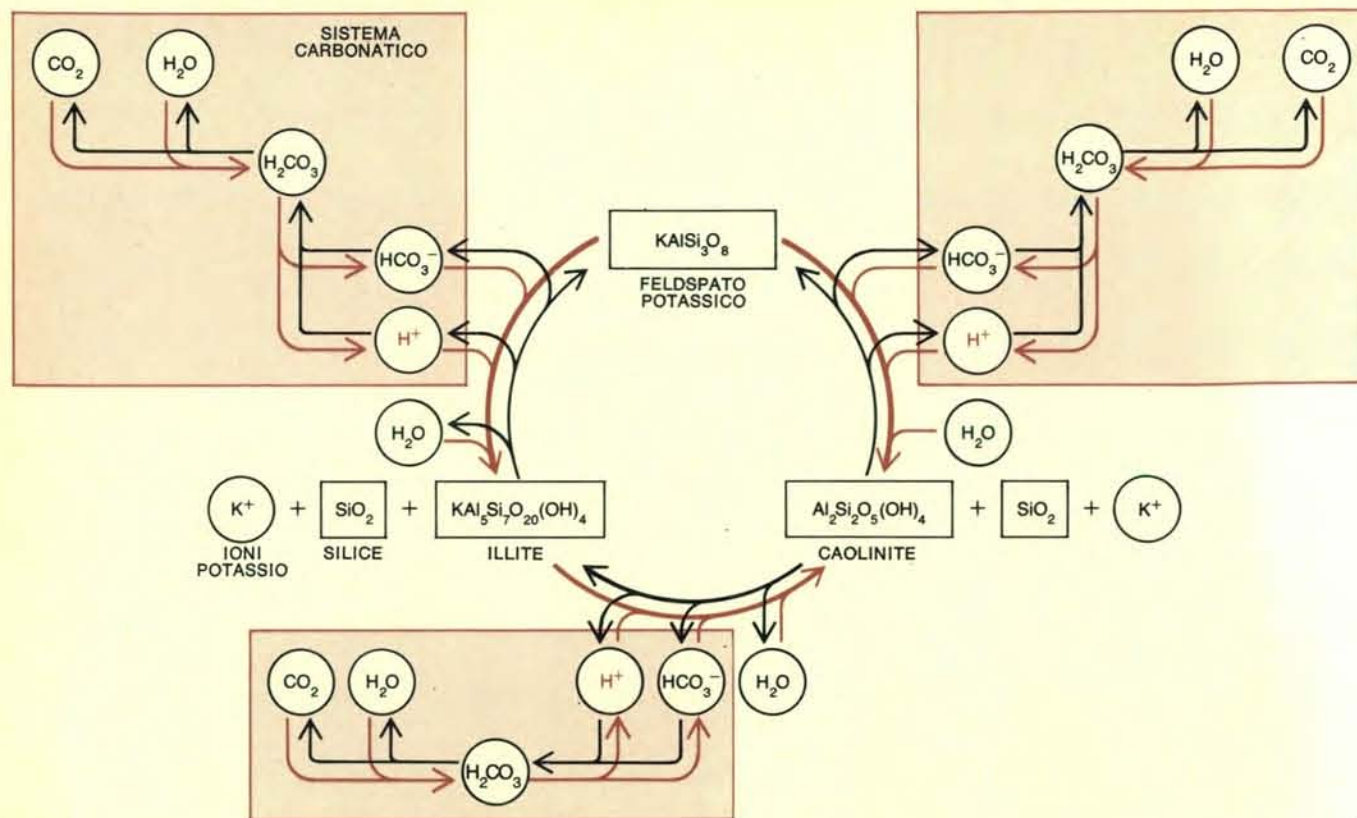
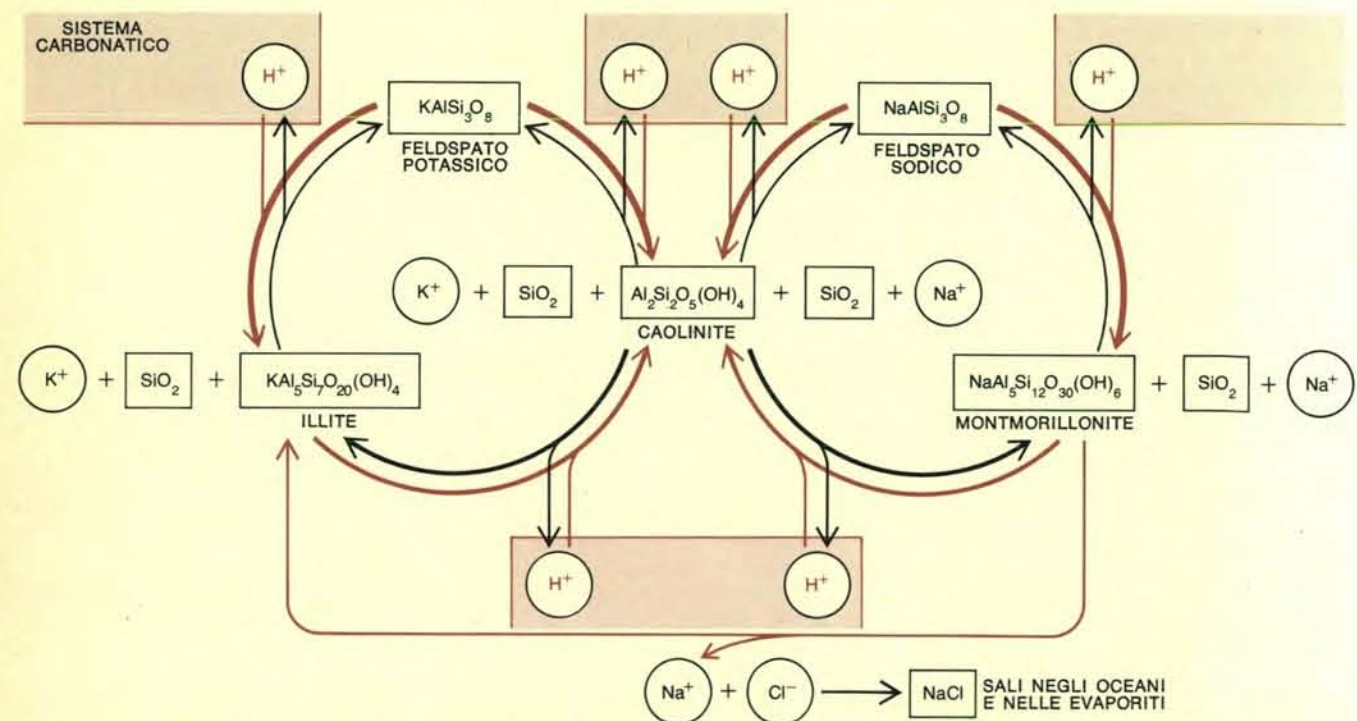
L'analisi ha inizio con la supposizione che vi sia equilibrio tra *input* e *output*. Si suppone inoltre che l'intero sistema rappresentato dalla crosta terrestre, dall'atmosfera e dagli oceani sia ora generalmente stabile e lo sia stato per lunghi periodi geologici nel passato. In breve, questo sistema gigantesco è considerato come se fosse in uno stato stazionario.

La supposizione dello stato stazionario è giustificata dal confronto dei materiali che entrano e lasciano gli oceani ogni anno. Fin dal 1952 T.F.W. Barth dell'Università di Oslo formulò l'ipotesi dello stato stazionario, valutò l'affluenza di solidi e di materiale sciolto negli oceani e venne fuori col concetto del « tempo di residenza di un elemento », che egli definì come la quantità totale di materiale presente nel sistema (nel suo caso gli oceani) divisa per la velocità dell'affluenza o per quella di efflusso, dato che nello stato stazionario queste due velocità sono uguali. Nel 1958 Edward D. Goldberg e Gustaf O.S. Arrhenius della Scripps Institution of Oceanography calcolarono i tempi di residenza per gli elementi negli oceani solo dai dati di efflusso, ossia la velocità alla quale questi componenti lasciano gli oceani sotto forma di sedimenti depositati sul fondo oceanico. Per molti elementi essi trovarono una notevole coincidenza con i tempi di residenza di Barth. Questa uguaglianza nelle velocità di affluenza e di efflusso è un argomento efficace se si suppone che gli oceani siano attualmente in uno stato stazionario e lo siano stati nel recente passato.

È meno facile dimostrare in questo modo che l'atmosfera sia esattamente in equilibrio rispetto ai suoi componenti o che la superficie solida della Terra sia in equilibrio stabile rispetto alla formazione delle montagne e all'erosione, ma si è incoraggiati a dimostrarlo dalla documentazione geologica. Le rocce antiche hanno lo stesso aspetto di quelle di recente formazione; inoltre i vari tipi di rocce compaiono grosso modo con le stesse percentuali caratteristiche in tutti i lunghi periodi di tempo riferiti alla scala geologica. Come prima approssimazione supponiamo che l'intero guscio esterno della Terra sia stato stabile almeno per la maggior parte del passato geologico noto, che copre circa 600 milioni di anni dall'inizio del Cambriano.

L'*input* primario di solidi entro questo sistema avviene attraverso la formazione delle montagne, che espone rocce profondamente sepolte all'attacco erosivo dell'atmosfera attraverso il meccanismo dell'alterazione, il processo fisico-chimico attraverso cui le rocce si disgregano. L'alterazione come processo chimico è la somma di parecchie reazioni che portano alla fissazione degli ioni idrogeno, alla rimozione dell'anidride carbonica e dell'ossigeno dall'atmosfera, e alla idratazione, carbonazione e ossidazione dei metalli ridotti che si trovano nei minerali ossidi e silicati di origine ignea. Una tipica reazione di alterazione coinvolge l'alterazione di un feldspato ricco in potassio ($2KAlSi_3O_8$), uno dei principali costituenti di molte rocce ignee, per formare caolinite ($Al_2Si_2O_5(OH)_4$), una argilla ricca in alluminio abbondante in molti suoli alterati: $2KAlSi_3O_8 + 2H^+ + 2HCO_3^- + H_2O \rightarrow Al_2Si_2O_5(OH)_4 + 2K^+ + 4SiO_2 + 2HCO_3^-$.

Gli ioni idrogeno necessari per questa reazione sono forniti dal sistema carbonatico, in cui l'anidride carbonica (CO_2) dell'atmosfera si combina con l'acqua di pioggia per formare acido carbonico (H_2CO_3), che poi si ionizza per formare ioni idrogeno (H^+) e ioni bicarbonato (HCO_3^-). Il risultato della reazione di alterazione è quello di asportare l'idrogeno dall'acido carbonico, e quindi di immobilizzare l'anidride carbonica come bicarbonato in soluzione e di fissare l'idrogeno e l'acqua nella caolinite solida. Contemporaneamente gli ioni potassio (K^+) e la silice (SiO_2) diventano parte del carico disciolto nelle acque superficiali, muovendosi eventualmente verso l'oceano, trasportati dalle acque. Si può quindi pensare ai silicati e ai feldspati inalterati come ad accettori di idrogeno e all'acido carbonico libero come a un donatore di idrogeno.

L'alterazione del calcare provoca quasi lo stesso risultato chimico pro-

Questa illustrazione mette in evidenza la natura ciclica dell'alterazione dei feldspati (*in colore*) e la ricostituzione dei silicati (*in nero*). Dai feldspati, che sono quasi interamente di origine ignea, l'alterazione forma varie specie di minerali argillosi, che in parte possono riformare piccole quantità di feldspati. La silice residua, che cioè non è stata sottratta per partecipare al ciclo delle argille, appare nei campioni geologici come calcedonio e opale. Gli ioni idrogeno necessari per le reazioni di alterazione sono forniti prevalentemente dal sistema carbonatico nel quale l'anidride carbonica dell'atmosfera si combina con l'acqua di pioggia per formare acido carbonico che poi ionizza producendo gli ioni idrogeno e gli ioni bicarbonato. Il particolare feldspato di questa tipica reazione di alterazione è il silicato di potassio e alluminio chiamato ortoclasio ($KAlSi_3O_8$).

Gli ioni potassio (K^+) e gli ioni sodio (Na^+) durante l'alterazione e la ricostituzione di rocce silicatiche si separano attraverso vari stadi. Originariamente il potassio e il sodio sono ben separati nei feldspati primari, ma essi si liberano in soluzione per alterazione tendendo eventualmente in definitiva ad andare verso l'oceano. Essi possono separarsi di nuovo nel corso della ricostituzione dei silicati, ma poiché la montmorillonite, che è ricca in sodio, gradualmente si trasforma in illite, che è invece ricca in potassio, il potassio si ferma nelle argille mentre il sodio va in soluzione nell'oceano o è immagazzinato nei depositi salini noti come evaporiti. Nella figura è messa in risalto, del sistema carbonatico, soltanto la funzione dello ione idrogeno.

dotto dall'alterazione dei silicati; gli ioni idrogeno sono utilizzati nella reazione come segue: $CaCO_3 + H^+ + HCO_3^- \rightarrow Ca^{++} + 2HCO_3^-$. Nella reazione di alterazione del calcare l'ione carbonato (CO_3^{--}) della calcite contenuta nel calcare è l'accettore di idrogeno. Pare che la maggior parte dell'alterazione chimica che avviene sulla superficie della Terra riguardi il calcare e che in confronto la quantità di silicati alterati sia piccola.

Un tipo diverso di reazione di alterazione consiste nella semplice ossidazione di un metallo ridotto, come il ferro. Si può dire dunque che la magnetite (Fe_3O_4) si combina con l'ossigeno per formare ematite (Fe_2O_3), ignorando per il momento la complessa catena di reazioni intermedie attraverso cui si ottiene questo risultato finale. La quantità di ossigeno gassoso richiesta per questa reazione è di ordini di grandezza inferiori alla quantità di ossigeno presente nell'atmosfera. La differenza non solo spiega la rapida ossidazione dei metalli delle rocce quando per la prima volta si trovano esposti in superficie, ma fa anche pensare che il livello di ossigeno nell'atmosfera non sia affatto controllato dall'equilibrio minerale tra gli stati ridotti e gli stati ossidati di metalli come il ferro.

La provenienza dell'anidride carbonica presente nell'atmosfera è stata oggetto di considerazioni da parte di W.W. Rubey nel 1951 in un saggio ormai classico sull'origine delle acque marine. Rubey ha dimostrato che il confronto tra la composizione delle rocce ignee e i sedimenti derivati dalle stesse per alterazione indicava una eccedenza di certi composti nelle rocce sedimentarie, negli oceani e nell'atmosfera. Egli chiamò queste sostanze – soprattutto anidride carbonica, acido cloridrico (HCl) e ione solforico (SO_2) – « volatili in eccesso ». Questi volatili in eccesso sarebbero prodotti dalla fuoriuscita di gas dall'interno della Terra attraverso i vulcani attivi, le sorgenti termali e i geyser. La maggior parte dell'anidride carbonica in eccesso, che si è accumulata a causa della fuoriuscita di gas, ha reagito col calcio sciolto proveniente dalle rocce ignee e metamorfiche alterate e ora è sepolto come carbonato di calcio ($CaCO_3$) nei calcari. L'acido cloridrico in eccesso si è accumulato nel mare o è sepolto in depositi fossili di sale.

Diamo ora uno sguardo complessivo al sistema e cerchiamo di trovare una spiegazione per la forza motrice che sta dietro alle reazioni chimiche di alterazione. Il modo più semplice per farlo è quello di considerare un modello in cui l'interno più profondo della Terra, le parti inferiori della crosta e le parti superiori del mantello sono legate in un ciclo controllato dalle differenze di temperatura e di pressione (*si veda l'illustrazione a pagina 35*). Le rocce fondono in luoghi profondi della crosta e nelle parti superiori del mantello a causa dell'alta temperatura e pressione che vi regnano; in questo senso l'interno della Terra assomiglia a un reattore caldo « non stagno ». La perdita riguarda le fasi fluide come l'acqua, l'anidride carbonica, l'acido cloridrico e alcuni elementi allo stato gassoso come lo zolfo, l'azoto e il fosforo: in breve tutti i volatili in eccesso che sono apparsi sulla superficie della Terra nel corso dei tempi geologici. La perdita si scarica in superficie dove può essere osservata e misurata nei pressi dei vulcani e delle sorgenti termali. Inoltre la geochimica dei sedimenti sepolti in profondità fornisce ampie prove che alcuni di questi volatili trovano sfogo anche negli strati inferiori delle rocce sedimentarie e metamorfiche profonde; questi volatili possono anche non raggiungere mai la superficie. I solidi degassificati che giacciono in profondità sono, a quelle alte temperature e pressioni, stabili; in parte per la perdita dei volatili e in parte per il raffreddamento, questi elementi chimici solidificano dando origine alle rocce ignee profonde della crosta terrestre. Infine, per azione dell'espansione dei fondi oceanici, della formazione delle dorsali medio-oceaniche e dei processi orogenetici, i prodotti solidi del reattore caldo vengono portati in superficie dove incontrano temperature e pressioni molto più basse.

Le fasi fluide che arrivano in superficie vanno nell'atmosfera, condensano o reagiscono chimicamente e in superficie vengono immagazzinati i prodotti della reazione chimica: argille e sedimenti alterati. L'idrogeno è il solo elemento che per la sua leggerezza può lasciare l'atmosfera terrestre e perdersi nello spazio. Anche l'elio sfugge, ma molto più lentamente dell'idrogeno. Tranne che per questi due elementi, si può considerare la superficie della Terra, compresa l'atmosfera, come un sistema chiuso: un reattore freddo in cui sono contenuti i prodotti del reattore caldo che si trova nelle profondità della Terra.

Man mano che le rocce ignee vengono in contatto con le fasi fluide raffreddate e condensate, ha luogo una reazione opposta a bassa temperatura tra i prodotti formatisi dalla reazione ad alta temperatura. I costituenti dominanti dei fluidi, soprattutto anidride carbonica e acqua, reagiscono con i silicati e gli ossidi delle rocce ignee a temperature di 500-1000 gradi centigradi più basse di quelle a cui si sono formati nell'interno della Terra. A causa della temperatura e della pressione più basse, si formano idrati e carbonati, che sono instabili a pressioni e temperature elevate.

Una trasformazione semplice che mostra questa reversibilità è il processo con cui la calcite, una forma di carbonato di calcio, reagisce col quarzo, una forma di silice, ad alte temperature durante il metamorfismo dando origine a wollastonite ($CaSiO_3$) più anidride carbonica, che sfugge verso la superficie. In superficie, quando la wollastonite è esposta all'aria che contiene anidride carbonica, la reazione procede in senso opposto e si formano due minerali tipicamente sedimentari: la calcite e il quarzo. Il metamorfismo e il riassorbimento nel mantello dei bordi delle zolle crostali nelle fosse oceaniche profonde costituiscono semplicemente il ritorno dei solidi al reattore caldo.

Il processo di alterazione può quindi essere considerato come un lato di un grande sistema chiuso – il mantello, la crosta, l'atmosfera e gli oceani – in cui c'è un gradiente di temperatura e di pressione. Qualsiasi cosa reagisca nella parte terminale calda, reagisce in direzione opposta quando si trova nella parte terminale fredda. L.G. Sillén dell'Istituto reale di tecnologia di Stoccolma considerava l'intero processo in un modo un po' diverso. Egli lo descriveva come una gigantesca reazione di neutralizzazione acido-base in cui i volatili che sfuggivano si combinavano con l'acqua per formare acidi, i quali a loro volta reagivano con i silicati e gli ossidi metallici ridotti delle rocce ignee e metamorfiche.

Consideriamo ora l'equilibrio tra oceano e atmosfera rispetto alla distribuzione dei prodotti dell'alterazione e dell'erosione, ossia le particelle solide in sospensione e il materiale in soluzione trasportato dai continenti al mare. L'alterazione consiste nella reazione con l'anidride carbonica; si ha come conseguenza la rimozione dell'anidride carbonica gassosa dall'atmosfera e la sua cattura in forma disciolta da parte degli ioni bicarbonato dell'acqua. Questi ioni sono trasportati dai fiumi e dagli oceani: gli ioni bicarbonato positivi sono equilibrati largamente dagli ioni positivi degli elementi disciolti nell'alterazione delle rocce. Ora, però, per mantenere i presupposti dell'ipotesi originaria – che ogni cosa è in un equilibrio dinamico e quindi la quantità dell'anidride carbonica nell'atmosfera è costante – l'anidride carbonica deve in qualche modo ritornare nell'atmosfera. Ma solo una piccola parte dell'anidride carbonica ri-

torna attraverso i gas vulcanici, perché l'alterazione è un processo molto più veloce rispetto all'accumulo di anidride carbonica di provenienza dal mantello attraverso miliardi di anni. Quindi la parte maggiore deve provenire dalla dissociazione di ioni bicarbonato in anidride carbonica gassosa. L'unico modo in cui il processo può compiersi è che l'anidride carbonica si combini con ioni idrogeno disponibili così che si possa nuovamente formare acido carbonico e quindi anidride carbonica e acqua.

Dove si può ottenere l'idrogeno necessario? Qualche elemento sulla funzione della sedimentazione a questo proposito si può trarre dalla stessa reazione chimica che descrive l'alterazione del calcare. Se si legge questa reazione in senso opposto – ossia come la reazione di ioni calcio positivi con ioni bicarbonato negativi per formare carbonato di calcio – essa descrive la precipitazione del calcare. Il fatto che molta parte di questa precipitazione abbia luogo biochimicamente nella produzione di rocce biocostruite calcaree (contenenti calcio) non dovrebbe rendere meno chiara la natura generale della reazione. L'aspetto importante della precipitazione del carbonato di calcio è che essa libera ioni idrogeno, che possono quindi combinarsi con ioni bicarbonato per formare anidride carbonica e acqua.

Poiché sappiamo che l'alterazione del calcare costituisce la parte maggiore di tutta l'alterazione chimica, e quindi il maggiore immobilizzatore di anidride carbonica, potrebbe essere utile calcolare l'equilibrio come se il calcare fosse l'unica roccia che controllasse l'anidride carbonica dell'atmosfera. In questa eventualità, per ogni mole, o grammomolecola, di anidride carbonica usata nell'alterazione e quindi rimossa dall'atmosfera, si dovrebbe disciogliere una mole di calcare depositato sul fondo oceanico. Fintanto che la precipitazione del calcare equilibra esattamente l'alterazione del calcare non ci sarebbero variazioni e il carbonato verrebbe semplicemente e costantemente riciclato attraverso l'atmosfera e gli oceani. Se la precipitazione di calcare, però, subisce un rallentamento, ne conseguirebbe una graduale riduzione dell'anidride carbonica nell'atmosfera.

Si potrebbe pensare che questa riduzione sia controbilanciata un po' dall'anidride carbonica proveniente dall'interno della Terra. La velocità di fuoriuscita dei gas è però, come abbiamo già detto, molto lenta e non potrebbe avere effetto lungo intervalli di tempo inferiori a centinaia di milioni di anni, mentre il tempo di residenza degli ioni carbonato nell'oceano è molto più breve, circa 100 000 anni. Il sistema invece ha cura di se stesso. Qualsiasi riduzione nella precipitazione di calcare causerebbe un accumulo nella concentrazione di ioni calcio e di ioni carbonato nell'oceano: il meccanismo di retroazione (*feedback*) negativa consisterebbe nel fatto che la solubilità del carbonato di calcio sarebbe tanto maggiorata nell'oceano che in definitiva la velocità di precipitazione comincerebbe ad aumentare e si ristabilirebbe l'equilibrio.

Nella figura sono riassunti i principali cicli geochimici, che operano al presente alla superficie della Terra, in relazione ai tre diversi serbatoi chimici: serbatoio idrogenionico, serbatoio oceanico e serbatoio sedimentario. I percorsi dei vari elementi « tracce » e dei composti sono indicati attraverso gli stadi di alterazione di calcari e silicati (che ricevono probabilmente un *input* dal serbatoio di idrogeno), la dissoluzione nel serbatoio oceanico e la deposizione (inclusa precipitazione e diagenesi) nel serbatoio sedimentario, che libera ioni idrogeno e molecole di anidride carbonica per il riciclo. Il serbatoio di idrogeno è rifornito prevalentemente dal sistema carbonatico e secondariamente dal sistema HCl (acido cloridrico).

Alcuni ioni idrogeno sono messi a disposizione dall'acido cloridrico che proviene dall'interno della Terra, ma in qualsiasi momento la quantità disponibile è piccola, come si può vedere confrontando la quantità di acido cloridrico con la quantità di anidride carbonica che esce in forma di gas. Tutto il cloro apparso sulla superficie della Terra nel corso dei tempi geologici, derivato in ultima analisi dall'acido cloridrico fuoriuscito come gas, si è accumulato o negli oceani o si è depositato nelle rocce sedimentarie così come nei depositi di sale o nelle acque saline. Una parte di esso è stata riciclata attraverso la crosta profonda e il mantello così che va sottratta dal bilancio. La quantità totale di cloro, però, è piccola in confronto alla quantità totale di anidride carbonica; dato che il tempo di residenza del cloro è molto più lungo di quello dell'anidride carbonica non ce ne può mai essere a sufficienza per produrre molto acido.

Un altro modo per ridare ioni idrogeno al sistema carbonatico, permettendo così il ritorno all'atmosfera di anidride carbonica, è quello di rovesciare le reazioni di alterazione e farle procedere nella direzione opposta. Questo processo è analogo a quello illustrato prima a proposito della precipitazione del calcare considerata come l'opposto della dissoluzione del calcare. Per i silicati, però, si tratta di un processo più complicato in cui gli ioni sodio, potassio, calcio e magnesio si combinano con la silice e i costituenti delle argille, come la caolinite, per formare nuovi silicati che sono più ricchi in silice e più poveri in idrogeno (e generalmente anche più poveri in acqua). Per esempio, la reazione con cui la caolinite si trasforma in un'altra argilla, l'illite, libera anidride carbonica.

Le trasformazioni di un minerale di argilla in un altro sono note come parte del complesso delle variazioni chimiche e fisiche postdeposizionali, chiamato diagenesi, attraverso cui i sedimenti di recente deposizione diventano rocce consolidate. Si possono considerare queste reazioni dei minerali argillosi come una ricostituzione di silicati, un processo che controbilancia l'alterazione, la quale consiste nella distruzione dei silicati (*si veda l'illustrazione in alto a pagina 36*). Il ciclo è generalmente incompleto; nonostante che i feldspati, i minerali più abbondanti della crosta terrestre, si alterino in larga misura per formare argille, solo una piccola parte di tutte le argille reagisce in senso opposto per ricostituire i feldspati nelle rocce sedimentarie. Nella parte sedimentaria della crosta solo le trasformazioni argilla--argilla sono effettivamente cicliche. Durante il metamorfismo e il riciclo dei sedimenti e delle rocce metamorfiche attraverso una fase ignea (palingenesi) il ciclo si completa fino allo stadio dei feldspati. L'intero ciclo quindi che coinvolge l'anidride carbonica è completo solo quando vi sono compresi processi che si ritiene generalmente non abbiano alcun rapporto con gli oceani e l'atmosfera. Una della implicazioni del ciclo per quanto riguarda la documentazione sedimentaria è che per ogni mole di feldspati alterati e non bilanciati da feldspati diagenetici precipitati nei sedimenti, 1,6 moli di silice sono tralasciate e trovano la loro destinazione in rocce come calcedonio o altre varietà di silice sedimentaria.

Anche la distribuzione del sodio e del potassio e le loro diverse abbondanze nelle argille fanno parte di questo ciclo generale; poiché questi metalli alcalini sono liberati in grandi quantità dall'alterazione dei feldspati, la loro sorgente primaria è nelle rocce ignee (*si veda l'illustrazione in basso a pagina 36*). Come funziona questo sottociclo? Sebbene la caolinite si formi originariamente come prodotto dell'alterazione dei due feldspati più abbondanti, uno ricco in potassio (l'ortoclasio e il microclino) e l'altro ricco in sodio (l'albite), le argille ricostituite incorporano i due elementi in maniera diversa. L'illite prende il potassio e la montmorillonite il sodio. A questo proposito la Terra agisce nel separare i due elementi quasi come un chimico analista. In ultima analisi il sodio trova una dimora diversa negli oceani e nei depositi di sale chiamati evaporiti; risulta evidente dalla documentazione rocciosa che la montmorillonite si trasforma gradualmente in illite nel corso di circa 100 milioni di anni. Sia il sodio sia il potassio si ricostituiscono in feldspati nel corso del metamorfismo o nella fusione ignea. Perché quindi un equilibrio di massa funzioni in modo che le rocce ignee mantengano il rapporto sodio-potassio, è chiaro che in questo processo vi deve essere una mescolanza appropriata nella crosta di materiali sedimentari, quali argille, evaporiti, acque saline e acqua marina.

Se la ricostituzione dei silicati è una parte importante del processo per mezzo del quale la concentrazione di ioni idrogeno nell'atmosfera e negli oceani è conservata rispetto all'anidride carbonica, alla silice, ai metalli e alle terre alcaline, si dovrebbe essere in grado di trovare dove e quando queste trasformazioni hanno luogo. Dove hanno le argille alterate la loro prima opportunità di inserirsi nell'appropriato ambiente geochimico? Questo luogo dovrebbe essere ricco di sodio, potassio calcio, magnesio e silice in un ambiente alcalino che assorbirebbe gli ioni idrogeno man mano che vengono prodotti dalle reazioni di ricostituzione. Il mescolarsi delle acque fluviali e delle argille trasportate in sospensione con l'acqua marina degli estauri e dei delta sembra a prima vista la supposizione migliore. La concentrazione della silice disciolta diminuirebbe se la ricostituzione avesse luogo e potesse così essere usata come traccia della reazione. Le ricerche condotte in questi ambienti mostrano però che, se vi è, la ricostituzione dei silicati è scarsa.

Supponendo che il rapido passaggio in un estuario rappresenti un tempo troppo breve perché abbia luogo una reazione così lenta, si potrebbero considerare i sedimenti dei fondi oceanici. Qui sembra vi siano le condizioni e i tempi opportuni. Alcuni anni fa partecipai a una serie di crociere oceanografiche durante le quali ho raccolto campioni di sedimenti da cui feci uscire l'acqua interstiziale proprio per poter trovare le prove delle reazioni di ricostituzione dalle variazioni nella composizione dell'acqua in contatto con le argille. Tuttavia sia l'esame dell'acqua sia l'esame mineralogico dei sedimenti non rivelarono nessuna delle reazioni postulate. In effetti, era evidente che accadeva in piccola misura l'opposto, cioè particelle di feldspato erose meccanicamente e trasportate con i sedimenti prevalentemente argillosi continuavano ad « alterarsi » sul fondo del mare arricchendo le acque interstiziali di potassio. Sembrava che i depositi di organismi che secernono silice (diatomee e radiolari) si dissolvessero per dare alte concentrazioni di silice disciolta e non si avevano prove che tale silice precipitasse e si incorporasse con le argille nelle reazioni di ricostituzione. Sono stato in grado di scoprire con il microscopio elettronico e

la diffrazione a raggi X (così come altri che hanno lavorato in questo campo) che vi sono bensì minuscole quantità di silicati ricostituiti nei sedimenti marini, ma sembra che queste quantità siano associate ad aree in cui il vulcanismo sottomarino ha avuto una funzione importante.

Il disappunto mi fece ritornare alle mie esperienze con le rocce sedimentarie antiche che si trovano sui continenti, dove gli studi mineralogici delle arenarie, degli scisti e dei calcari hanno reso chiaramente evidente che le trasformazioni dei silicati hanno effettivamente avuto luogo. Grandi quantità di argille ricche in potassio (come l'illite) e ricche in magnesio (come la clorite), di feldspati diagenetici e di silice libera in forma di ortoclasio e di quarzo diagenetico sono caratteristiche di sedimenti trovati nei bacini sedimentari profondi. Qui il passaggio del tempo, la profondità e la reazione con le acque interstiziali di diversa composizione hanno dato luogo alla formazione di raccolte diagenetiche di minerali. Man mano che i sedimenti affondano in questi bacini, gradualmente si portano in regioni con temperatura e pressione più elevate che accelerano le reazioni in direzione della ricostituzione. Questa tendenza implica che vi può essere una costante, ma straordinariamente lenta, esalazione dai continenti di anidride carbonica verso l'alto come conseguenza della diagenesi delle rocce sedimentarie e delle acque interstiziali a esse associate.

Se si cercano attentamente i luoghi di massima diagenesi, in termini sia di quantità di rocce coinvolte sia di intensità del processo, è evidente che i bordi continentali sono il luogo di massimo accumulo di sedimenti, sia nelle geosinclinali marginali sia nelle piattaforme e nelle scarpate continentali. È in queste stesse fasce che hanno avuto luogo i principali episodi regionali di orogenesi e di metamorfismo delle rocce. Sempre qui vi sono le zone di subduzione nelle quali spessi accumuli di sedimenti oceanici sono trasportati, dalle zolle litosferiche che si muovono verso il basso, e vengono piegati e spinti sempre più in profondità. Le reazioni chimiche diagenetiche e metamorfiche che avvengono nei bordi continentali forniscono gli ioni idrogeno che in ultima analisi sono responsabili del continuo rifornimento degli oceani e dell'atmosfera.

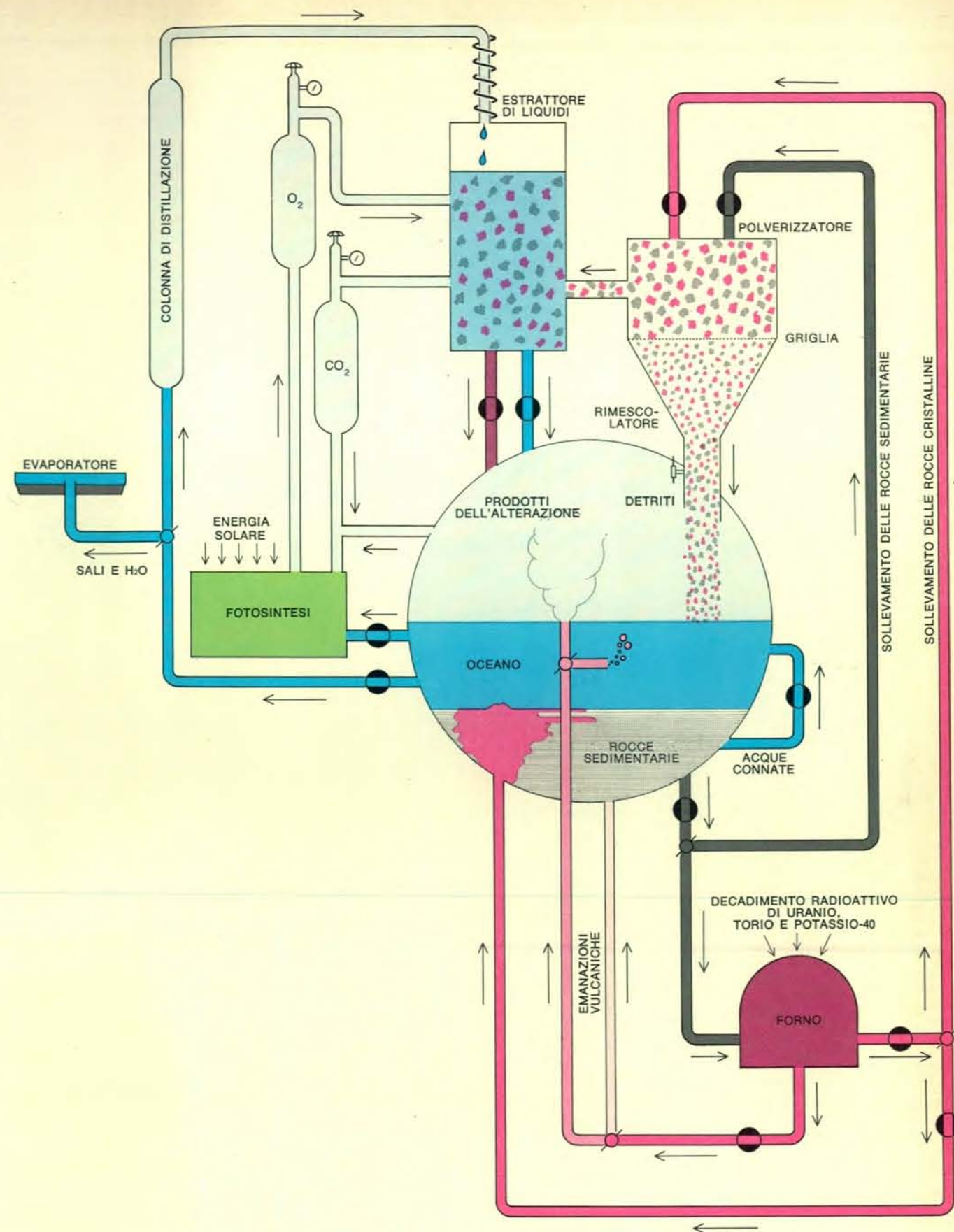
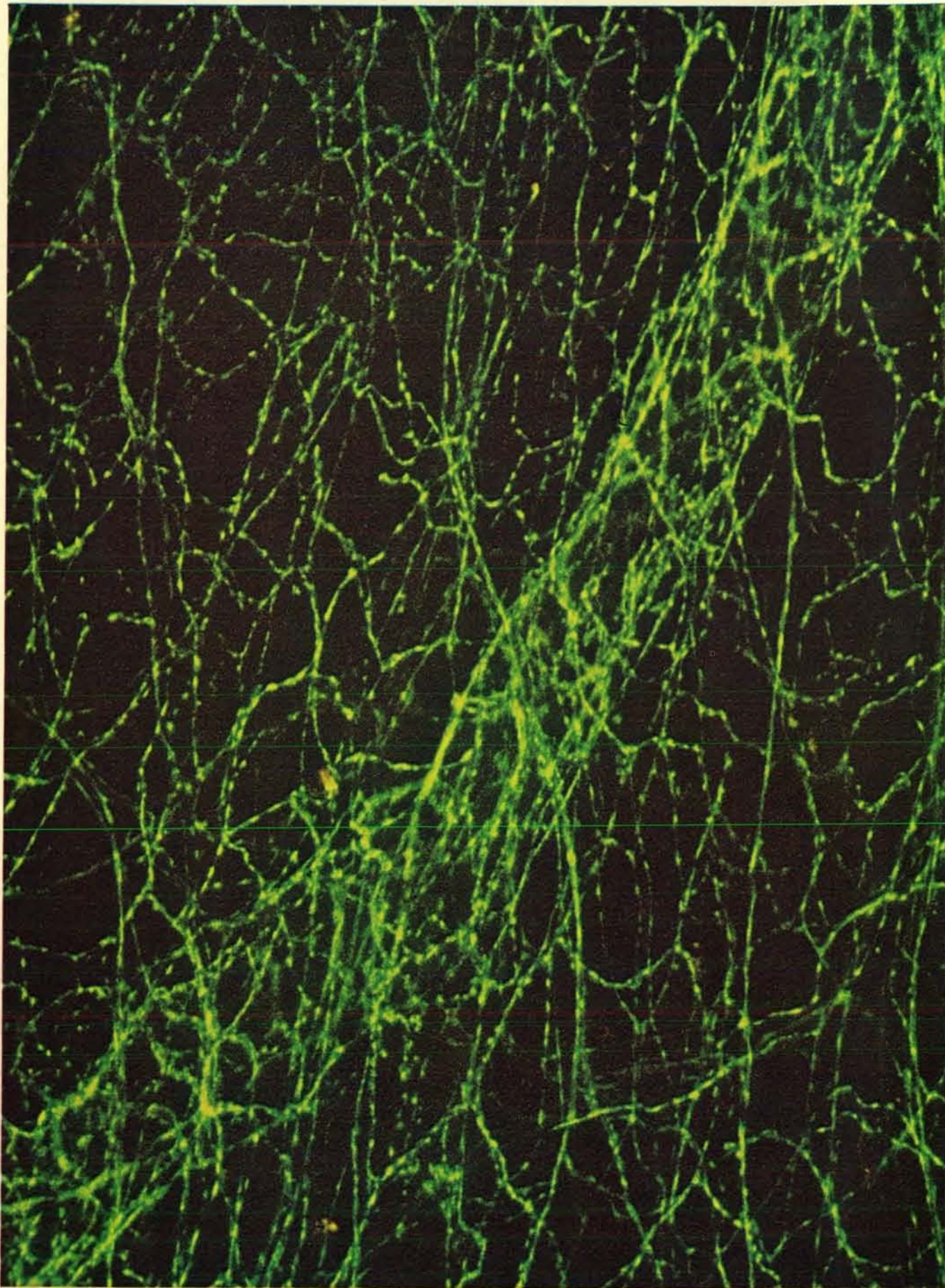
Attualmente si stanno raccogliendo prove per mezzo di profonde perforazioni sottomarine sull'esistenza di una circolazione di acque profonde nelle piattaforme continentali e probabilmente anche nelle scarpate. È opinione diffusa a proposito dei movimenti delle acque sotterranee nei sedimenti delle pianure costiere che le acque si diffondano negli oceani durante tempi molto lunghi. Gli ioni idrogeno, quindi, liberati sia dalla ricostituzione dei silicati sia dalla precipitazione dei carbonati arrivano negli oceani per rifornirli: gli accumuli sedimentari continentali di spessore relativamente piccolo, lontani dagli oceani, non sono collegati al sistema nel suo insieme tranne che in tempi molto lunghi; peraltro non sono volumetricamente importanti per il rifornimento del sistema.

L'intero processo può essere riassunto in un diagramma di funzionamento relativo all'alterazione, alla sedimentazione e alla ricostituzione rispetto al serbatoio di ioni idrogeno, al serbatoio oceanico e a quello sedimentario (*si veda l'illustrazione a pagina 38*). Un altro modo di illustrare questo sistema è quello di paragonare il sistema geochimico terrestre a un gigantesco impianto chimico: un concetto suggerito in una conferenza alcuni anni fa da Kenneth S. Deffeyes, ora alla Princeton University. Caratteristica stimolante di questo ipotetico impianto è la molteplicità di valvole e di interruttori che controllano il funzionamento del sistema e la velocità alla quale il sistema risponde a perturbazioni, oscillazioni o squilibri irregolari se vi sono violenti spostamenti in qualcuno degli interruttori e delle valvole (*si veda l'illustrazione nella pagina a fronte*). I regolatori di gas dei serbatoi di anidride carbonica e di ossigeno sono di particolare importanza per l'uomo, e in realtà per tutta la vita respiratoria.

Il regolatore di ossigeno è governato soprattutto dalla fotosintesi e solo secondariamente dall'alterazione. Se una Terra priva di vegetazione fosse abbandonata a se stessa per un tempo geologicamente lungo, l'ossigeno dell'atmosfera verrebbe gradualmente usato per ossidare i metalli ridotti portati in superficie dall'orogenesi e dall'attività ignea. Il regolatore dell'anidride carbonica è probabilmente molto più controllato dall'alterazione che non dalla fotosintesi. Il diagramma è ovviamente incompleto poiché ho omessso gli importanti sottocicli dell'azoto e del fosforo due elementi fortemente legati all'origine della vita e ai processi biologici in generale. Anche se i particolari differiscono, però, il modello operativo generale è lo stesso per tutti gli elementi reagenti.

Questa analisi ci conduce a ricercare i possibili diversi stati del sistema che si sono avuti nel passato tramite l'esame dei tipi e delle quantità di sedimenti geologici relativamente a quelli attuali. Mentre oggi il sottociclo del calcare nella alterazione e nella precipitazione è molto più importante che non il sistema dei silicati nel fornire l'anidride carbonica, può darsi che nel corso del tempo geologico non fosse così. Poiché sia la formazione del calcare sia la ricostituzione dei silicati forniscono ioni idrogeno, la loro somma dovrebbe essere costante per mantenere nell'atmosfera e negli oceani la stessa quantità di anidride carbonica. Di conseguenza, nei tempi passati, in cui vi era una minor produzione di calcare e minore alterazione, si sarebbero dovuti produrre accumuli di rocce con una quantità di silicati ricostituiti maggiore in confronto al calcare di quanto non avvenga nei depositi attuali o recenti.

Un'altra conseguenza dello stato stazionario è che la deposizione di sali e la ricostituzione di silicati dovrebbero dare una somma complessiva costante al fine di conservare costante il contenuto di metalli alcalini degli oceani. Vi sono molti altri equilibri che riflettono eventuali variazioni nel funzionamento passato del sistema; tutto dipende da una analisi particolareggiata della distribuzione mondiale dei tipi di rocce in un dato periodo.

Il passato della Terra è affare dei geologi, ma il suo futuro riguarda tutti. Per conoscere come l'enorme immissione di anidride carbonica nell'atmosfera avutasi negli ultimi cento anni per la combustione di combustibili fossili influenzerà l'intero sistema, abbiamo bisogno di sapere con quale rapidità reagisce il sistema e quali sono le conseguenze immediate. Anche se le riserve di ossigeno nell'atmosfera sono così ingenti da non doverci preoccupare per i prossimi millenni, l'equilibrio della produzione di ossigeno fotosintetico e della richiesta di ossigeno da parte degli organismi, della materia organica in decomposizione e dei metalli ridotti può essere modificato in maniera sensibile dall'attività dell'uomo. La produzione di ossigeno per fotosintesi può essere influenzata in qualche modo da forme di inquinamento che agiscano sulla produttività delle alghe marine oppure da tagli estesi di foreste tropicali come quelle brasiliane. Quanta importanza può avere la velocità con cui l'ossigeno viene usato dai metalli ridotti delle rocce allorché la velocità dell'erosione viene drammaticamente aumentata dalle attività dell'uomo sul paesaggio? Qual è la velocità con cui l'ossigeno viene consumato da tutti i rottami di ferro rugginosi sparsi nel mondo? Finché non comprendiamo in modo più esauriente il funzionamento di questo vasto sistema non sapremo fino a che punto lo stiamo alterando.

Il sistema geochimico della superficie terrestre è qui rappresentato, per analogia, come un gigantesco impianto chimico. La sorgente primaria di energia per l'impianto è l'energia solare; un'altra sorgente di energia proviene dal decadimento di elementi radioattivi nel «forno», che rappresenta l'interno della Terra. Una caratteristica appariscente di un tale impianto ipotetico è il gran numero di valvole e di interruttori, che controllano il funzionamento del sistema e la velocità dei processi.

Le terminazioni nervose dei nervi simpatici provenienti dall'iride di occhio di ratto emettono fluorescenza verde dopo trattamento con formaldeide, dimostrando che contengono noradrenalina. Le terminazioni, in cui si trovano i bottoni dove la noradrenalina viene immagazzinata, sono ingrandite 2400 volte in questa microfotografia a fluorescenza eseguita da D. Jacobowitz.

I neuromediatori

Queste sostanze chimiche liberate dalle terminazioni nervose sono i messaggeri con cui le cellule nervose comunicano; esse mediano varie funzioni tra cui la contrazione muscolare e il controllo del comportamento

di Julius Axelrod

Nel 1901 il noto fisiologo inglese J.N. Langley rivelò che iniettando in animali un estratto di ghiandola surrenale si stimolavano i tessuti innervati dalle fibre del sistema nervoso simpatico provocando l'accelerazione della frequenza cardiaca, l'aumento della pressione sanguigna e la contrazione della muscolatura liscia. Appena tre anni prima, John J. Abel, della Johns Hopkins University, aveva isolato l'adrenalina dalle ghiandole surrenali. Così, l'osservazione compiuta da Langley spinse T.R. Elliott, suo studente all'Università di Cambridge, a iniettare adrenalina in animali da esperimento. Poté in tal modo rilevare che l'ormone, somministrato sotto forma di estratto non purificato, induceva in taluni organi una reazione simile a quella evocata dalla stimolazione elettrica dei nervi simpatici. Questo ricercatore formulò l'ipotesi che l'adrenalina potesse venire liberata dalle fibre del sistema nervoso simpatico provocando una reazione nelle cellule dei muscoli all'altezza delle giunzioni neuromuscolari. Elliott enunciò così, per primo, il concetto di comunicazione tra neuroni attraverso mediatori chimici. Un neuromediatore è una sostanza chimica liberata dalle terminazioni delle fibre nervose: esso viene identificato da un recettore posto sulla superficie di una cellula nervosa postsinaptica o di un'altra cellula eccitabile postgiunzionale ed è in grado sia di stimolare, sia di inibire la seconda cellula. È oggi dimostrato che esistono differenti neuromediatori capaci di agire su tessuti e processi fisiologici diversi. I neuromediatori fanno accelerare o ridurre la frequenza cardiaca, contrarre o rilasciare i muscoli. Inducono le ghiandole a sintetizzare enzimi che producono ormoni o li fanno secernere. Rappresentano inoltre i mezzi con cui il cervello controlla i movimenti e le variazioni affettive e comportamentali.

L'ipotesi di Elliott sulla trasmissione chimica riuscì ad affermarsi molto lentamente. Langley, che disprezzava le teorie di qualsiasi tipo, convinse il suo allievo a non approfondire ulteriormente le sue ipotesi fino a quando non avesse conseguito un maggior numero di dati sperimentali. Ciò richiese molto tempo: la prima prova definitiva della trasmissione neurochimica fu infatti ottenuta nel 1921 da Otto Loewi, che allora lavorava all'Università di Graz, in Austria, con un esperimento risolutivo che troncò qualsiasi discussione. Loewi pose due cuori di rana in un recipiente dove potevano continuare a contrarsi. Quando il ricercatore stimolava il nervo vago (un nervo del sistema parasimpatico che rallenta la frequenza cardiaca) del primo cuore, anche la frequenza del secondo cuore risultava rallentata, dimostrando in tal modo che qualche sostanza veniva liberata dal nervo stimolato, veniva trasportata dal liquido e agiva sul cuore perfuso. Questa sostanza fu poi identificata da Sir Henry Dale come acetilcolina, uno dei primi neuromediatori conosciuti. In un esperimento analogo la stimolazione dei nervi simpatici (che accelerano la frequenza cardiaca) di un cuore di rana aveva avuto come risultato quello di accrescere il numero dei battiti di un altro cuore perfuso dalla stessa soluzione, ma non stimolato.

I mediatori

Una sostanza chimica per essere classificata come neuromediatore deve rispondere a un certo numero di criteri: le fibre nervose devono possedere gli enzimi capaci di produrre la sostanza e di liberarle, se stimolate. Il neuromediatore, a sua volta deve essere in grado di reagire con un recettore specifico, localizzato sulla cellula postsinaptica e produrre una risposta di tipo biologico. In base a questi criteri due composti sono ora riconosciuti come neuromediatori: l'acetilcolina e la noradrenalina. Le fibre nervose che le contengono vengono denominate rispettivamente colinergiche e adrenergiche. Esiste poi un certo numero di sostanze chimiche contenute nelle fibre nervose che presentano molti dei criteri richiesti, ma non tutti. Appartengono a questo gruppo di supposti mediatori la dopammina, l'adrenalina, la serotonina, l'octopammina, l'istamina, l'acido gamma-amminobutirrico (GABA), l'acido glutammico, l'acido aspartico e la glicina.

In questo articolo prenderemo in considerazione una classe di neuromediatori, le catecolammine, poiché i dati relativi a questi composti sono tra i più completi e molti dei principi che controllano la loro attività sembrano regolare quelle dei mediatori in generale.

Le catecolammine comprendono la noradrenalina (chiamata anche norepinefrina), la dopammina e l'adrenalina (o epinefrina). Possiedono in comune una struttura chimica con un anello benzenico che porta due gruppi idrossilici adiacenti e una catena laterale costituita da etilammina. La noradrenalina è presente nei nervi periferici, nel cervello, midollo spinale e nella midollare surrenale. Nei tessuti periferici e nel cervello la noradrenalina agisce come un neuromediatore, cioè esplica la maggior parte dei suoi effetti localmente, sulle cellule postsinaptiche. Nella midollare surrenale si comporta invece come un ormone: viene immessa nel circolo sanguigno e agisce su organi bersaglio situati a distanza. La dopammina, che una volta era considerata un semplice prodotto intermedio nella sintesi della noradrenalina e dell'adre-

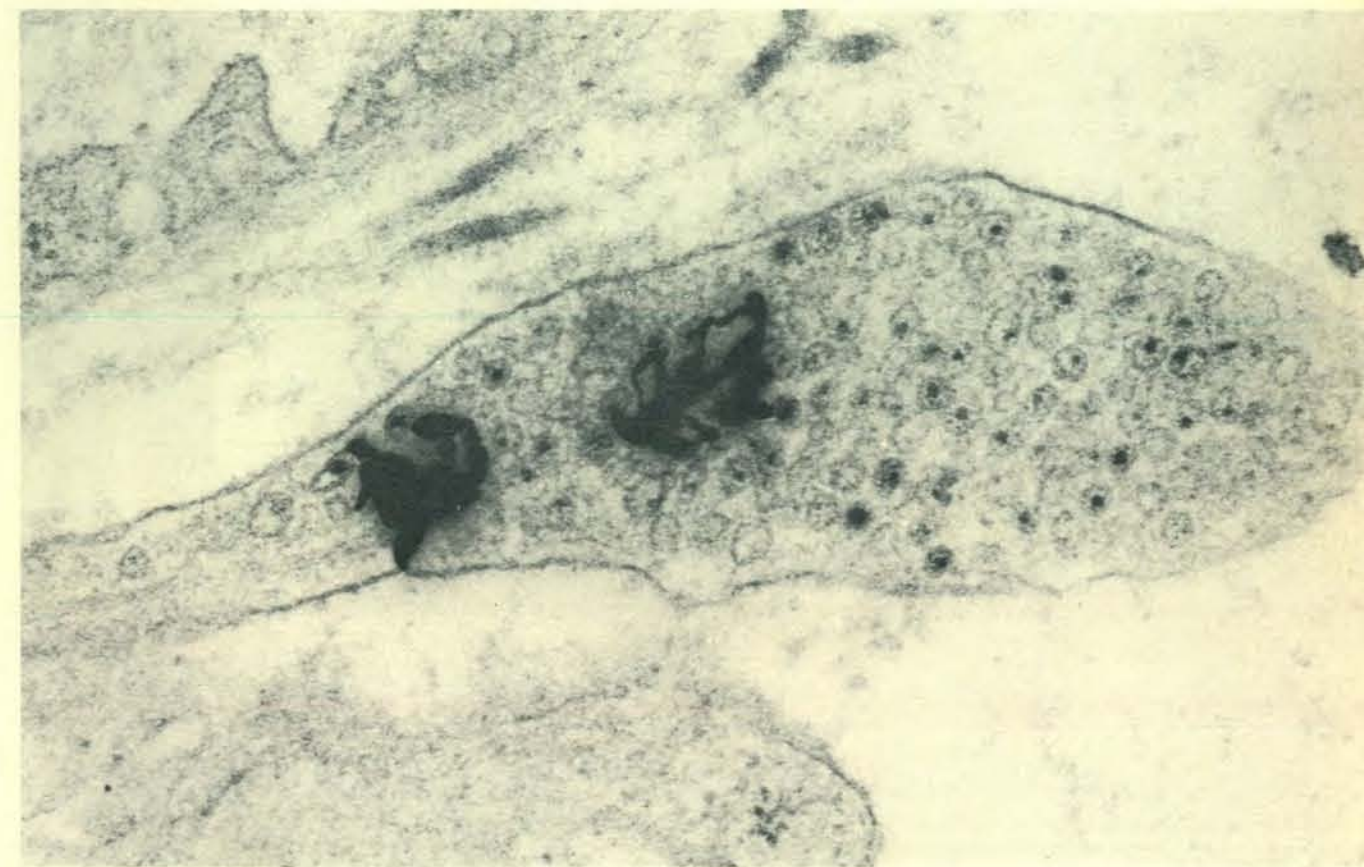
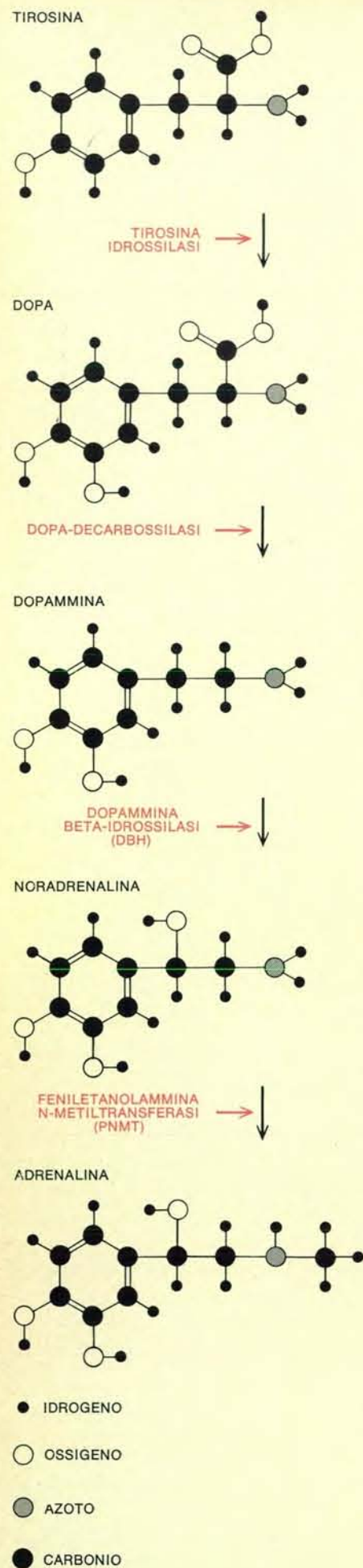
nalina, si comporta invece da neuromediatore nel cervello dove agisce sui nervi che condizionano l'attività motoria e il comportamento. La terza catecolammina, l'adrenalina, è presente in quantità elevate nella midollare surrenale. Viene liberata nel sangue in condizioni di stress come paura, collera o altro e agisce come un ormone su taluni organi, compresi il cuore, il fegato e l'intestino. Qualche anno fa, si è potuto dimostrare che anche l'adrenalina è probabilmente un neuromediatore giacché si trova nelle fibre nervose e nel cervello. La tecnica sviluppata un decennio fa in Svezia permette di visualizzare direttamente nei neuroni le catecolammine mediante la fluorescenza che esse emettono dopo trattamento con vapori di formaldeide. Il neurone possiede un corpo cellulare (pirenoforo) e un lungo assone, o fibra principale (detta anche neurasse) che si ramifica in un numero elevato di terminazioni. Ciascuna terminazione nervosa è dotata di varicosità, o espansioni (definite più propriamente bottoni sinaptici) che rassomigliano a perle infilate su un filo, cosicché ciascun neurone simpatico può innervare migliaia di altre cellule chiamate cellule effettrici. Nel 1960 Georg Hertting, Gordon Whitby e io siamo riusciti a dimostrare che la noradrenalina radioattiva (noradrenalina in cui taluni atomi di idrogeno sono sostituiti da atomi di trizio, l'isotopo radioattivo dell'idrogeno) è assunta selettivamente e trattenuta nelle fibre nervose del sistema simpatico. Nel laboratorio dell'autore, al National Institute of Mental Health, si è tentato di verificare se il neuromediatore viene immagazzinato all'interno del neurone. Le microfotografie al microscopio elettronico dei bottoni dei neuroni simpatici hanno rivelato un elevato numero di vescicole contenenti granuli elettronicamente densi. Quando la pellicola fotografica veniva esposta ai tessuti prelevati da ratti in cui era stata iniettata noradrenalina marcata, i granuli d'argento sviluppati dalla radiazione proveniente dagli atomi di idrogeno radiattivo risultavano rigorosamente localizzati sopra le vescicole contenenti i granuli densi. Ciò dimostrava che proprio in queste vescicole veniva immagazzinata la noradrenalina.

Sintesi e liberazione

Il processo che porta alla sintesi dei mediatori catecolamminici inizia nel corpo cellulare che è la sede dove si elaborano i quattro enzimi necessari per la loro formazione: tirosinaidrossilasi, DOPA-decarbossilasi, DOPA-ammina betaidrossilasi (DBH), feniletanolammina N-metiltransferrasi (PNMT). Gli enzimi sintetizzati nel corpo cellulare sono trasportati lungo l'assone attraverso un processo naturale di flusso (flusso assoplasmico) sino alle terminazioni nervose dove viene effettuata la sintesi delle catecolammine.

La liberazione del neuromediatore dalle terminazioni nervose implica una complessa serie di eventi. Quando una fibra nervosa viene stimolata, la permeabilità della sua membrana varia temporaneamente facendo penetrare gli ioni sodio, mantenuti nell'ambiente esterno in condizioni di riposo con un meccanismo attivo, e facendo fuoriuscire gli ioni potassio che si trovano all'interno della cellula. Questo passaggio di ioni provoca una variazione del potenziale di menbrana (depolarizzazione) che può propagarsi con un meccanismo « ondulatorio » attraverso i nodi privi di guaina mielinica e rivestiti solo da una sottile membrana che si trovano lungo l'assone. La depolarizzazione, e quindi la variazione del potenziale di membrana, può quindi trasmettersi da un nodo all'altro fino a raggiungere ed eccitare una sinapsi o una placca motrice. Come dimostrò per primo Bernhard Katz, dell'University College di Londra, per l'acetilcolina, la depolarizzazione provoca la liberazione di un quanto – o pacchetto – di mediatore chimico dalla terminazione nervosa nella fessura sinaptica. Dati biochimici ottenuti di recente nel laboratorio dell'autore e in altri dimostrano che la noradrenalina è liberata dalle fibre nervose pressappoco nello stesso modo. Le vescicole nelle terminazioni contengono non solo noradrenalina, ma anche l'enzima DOPA-ammina beta-idrossilasi (DBH) che trasforma la dopammina in noradrenalina. Quando la fibra simpatica è stimolata elettricamente, la noradrenalina e l'enzima sono liberati all'incirca nelle stesse quantità in cui sono presenti nelle vescicole. Questo processo sembrerebbe avvenire attraverso la fusione della vescicola con la membrana esterna della fibra nervosa, seguita dalla formazione di un'apertura (poro) sufficientemente ampia per consentire alle molecole di noradrenalina di essere espulse insieme con quelle, molto più voluminose, dell'enzima. Un meccanismo di liberazione di questo tipo è denominato esocitosi. Le singole fasi attraverso cui la vescicola si fonde con la membrana del neurone e forma un poro per liberare il suo contenuto solubile non sono ancora state specificate, così come il successivo destino della vescicola. Sono note talune condizioni che inibiscono la liberazione di noradrenalina e di DBH: una è costituita dalla presenza di « vinblastina » un composto che distrugge nei neu-

Il neurone del sistema nervoso simpatico, o cellula nervosa (*a*) consiste di un corpo cellulare, di un lungo assone e numerose terminazioni nervose costellate di bottoni sinaptici. Gli enzimi implicati nella sintesi dei mediatori sono elaborati nel nucleo e trasportati lungo l'assone sino ai bottoni dove vengono sintetizzati e immagazzinati i mediatori. In *b*, un bottone e la sua sinapsi con l'altra cellula, ingranditi. Il mediatore, (*punti colorati*) è immagazzinato nelle vescicole. Quando la fibra nervosa è eccitata, le vescicole si avvicinano alla membrana cellulare, fondendosi con essa e liberando il mediatore nella fessura sinaptica. Il mediatore raggiunge i recettori che riconoscono solo lui e nessun altro composto chimico (*forme nere*) presente nella fessura.

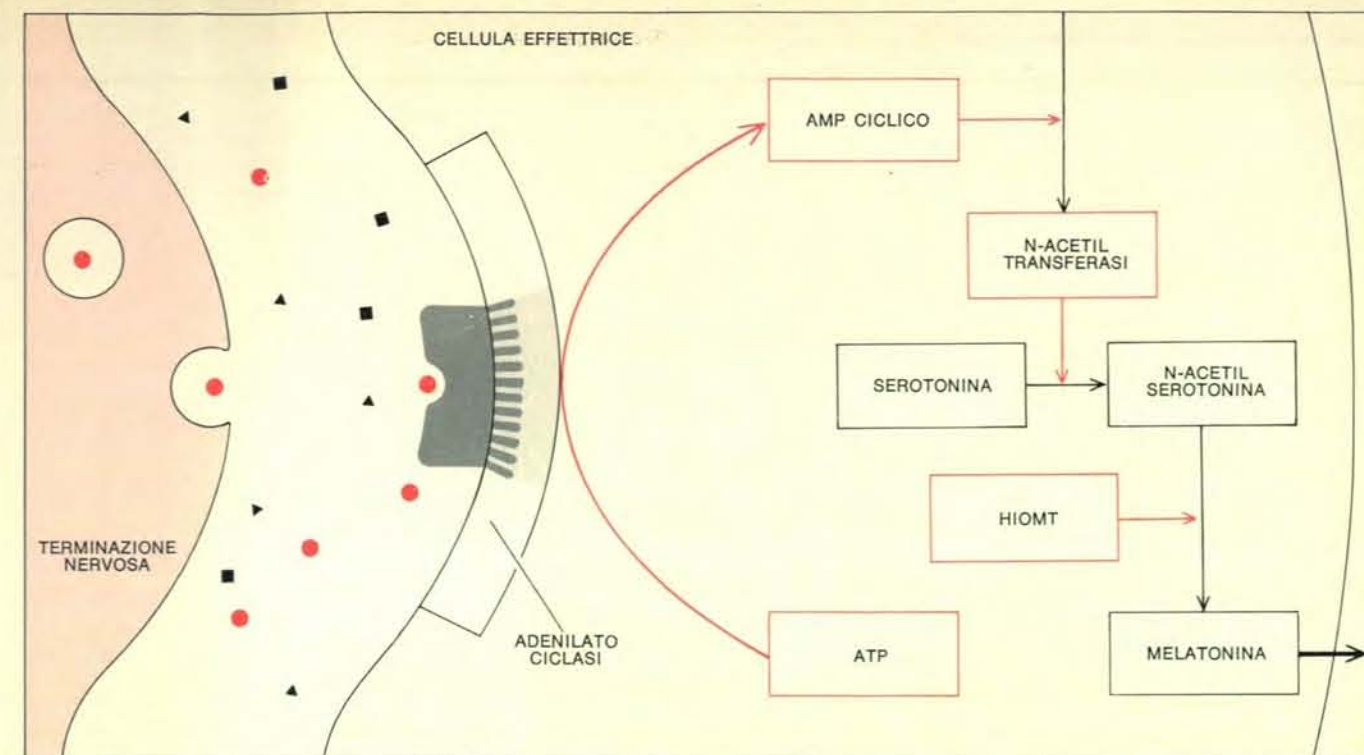
La microfotografia riproduce bottoni sinaptici posti in terminazioni nervose contenenti noradrenalina provenienti da epifisi di ratto e ingranditi 90 000 volte. I bottoni contengono vescicole, molte delle quali presentano granuli elettronicamente densi.

L'autoradiografia dimostra che la noradrenalina radioattiva è localizzata nelle vescicole. Un campione di tessuto epifisario è stato prelevato da ratti a cui era stata somministrata noradrenalina marcata. Questa ha fatto sviluppare i granuli d'argento di una pellicola fotografica collocata sopra il campione. I granuli erano localizzati sulle vescicole, dove si accumula il mediatore.

roni le strutture note come neurotubuli. Un'altra è rappresentata dalla presenza di « citocolasina-beta », una sostanza che altera la funzione del sistema di filamenti contrattili delle cellule. Una terza condizione è, infine, l'assenza di calcio. Questi dati dimostrerebbero che le lunghe strutture proteiche tubuliformi possono orientare le vescicole verso un sito localizzato sulla membrana neuronale in cui si verifica la liberazione del mediatore. È ben noto che i microfilamenti possono essere attivati dal calcio e contrarsi non solo nelle cellule nervose, ma anche in altre cellule, come quelle muscolari. È pertanto possibile che la depolarizzazione induca il calcio ad attivare un filamento contrattile nella membrana del neurone che di conseguenza si contrae provocando un'apertura sufficientemente ampia perché sia liberato il contenuto solubile della vescicola.

L'osservazione che la DBH veniva liberata dalle fibre nervose indusse Richard Weinshilboum a formulare l'ipotesi, che l'enzima potesse diffondersi attraverso il circolo sanguigno; fu elaborato un test altamente differenziato e si trovò che esso era effettivamente presente nel sangue. Insieme ad altri ricercatori si incominciò a determinare il valore dell'enzima (che è rilevabile specificatamente nei nervi simpatici) in diversi stati morbosi. Tale valore risulta ridotto in quella turba ereditaria del sistema autonomo nota come disautonomia familiare e nella sindrome di Down (mongolismo) ed elevato nella torsione distonica (una forma neurologica che implica una spasticità muscolare), nel neuroblastoma (una neoplasia del tessuto nervoso) e in talune forme di ipertensione. I dati osservati dimostrano che in ciascuna di queste malattie esistono anormalità funzionali del sistema nervoso simpatico.

Azione e inattivazione

Quando il neuromediatore viene liberato diffonde attraverso la fessura sinaptica fra la fibra nervosa terminale e le cellule adiacenti. La capacità della cellula effettrice più vicina di reagire con il mediatore dipende poi dalla presenza di un recettore posto sulla superficie cellulare postsinaptica che riconosce e si combina con il neuromediatore. Dopo l'interazione fra recettore e mediatore, viene iniziata una serie di eventi che inducono la cellula effettrice a svolgere la sua funzione specifica. Alcune di queste reazioni si attuano in maniera rapida (in una frazione di secondo), come nel caso della propagazione dell'impulso nervoso attraverso la sinapsi, altre si manifestano con più lentezza, in minuti, talvolta in ore, come nella sintesi degli enzimi intracellulari. Esistono due recettori che riconoscono la noradrenalina, i recettori alfa e beta adrenergici, e uno per la dopammina. Questi recettori si possono reciprocamente differenziare tramite la reazione specifica che ciascuno provoca e la capacità posseduta da determinati farmaci di bloccare questa reattività.

La produzione dei mediatori catecolamminici è associata a quattro enzimi situati nelle terminazioni dei nervi simpatici e nelle capsule surrenali. La tirosina viene trasformata nell'intermediario DOPA. Lo spostamento di un gruppo carbossilico (COOH) forma dopammina, mediatore e precursore della noradrenalina, un altro mediatore. Nelle capsule surrenali il processo continua con l'aggiunta di un gruppo metilico (CH_3) per formare adrenalina.

I recettori beta adrenergici stimolano le cellule effettrici mediante l'adenosina-3',5'-monofosfato o AMP ciclico, il « secondo messaggero universale » che interagisce con gli ormoni e molte attività cellulari da loro indotte (si veda l'articolo *AMP ciclico* di Ira Pastan, in « Le Scienze », novembre 1972). I ricercatori hanno potuto specificare numerose fasi dell'attivazione del recettore da parte della noradrenalina studiando l'interazione fra questa e le cellule adipose, epatiche ed epifisarie. Si è potuto riscontrare che l'epifisi, produttrice di un ormone chiamato melatonina, in grado di inibire l'attività delle ghiandole sessuali, è particolarmente adatta perché può essere facilmente innervata con fibre contenenti noradrenalina. La melatonina viene sintetizzata attraverso varie fasi, una di queste, la trasformazione della serotonina in N-acetilserotonina, è catalizzata dall'enzima serotonina N-acetiltransferasi. È questa sintesi enzimatica che è controllata dal recettore beta adrenergico. Quando la noradrenalina è liberata dalle fibre nervose che innervano l'epifisi, essa interagisce con i recettori beta adrenergici sul lato esterno della membrana di una cellula epifisaria. Allorché il recettore è saturato dalla noradrenalina, si attiva l'enzima adenilciclasi, sul lato interno della membrana cellulare. L'adenilciclasi trasforma poi l'ATP cellulare, ricco di energia, in AMP ciclico, che, a sua volta, stimola la sintesi della serotonina N-acetiltransferasi (*si veda l'illustrazione nella pagina a fronte*). Questa complessa serie di eventi può essere alterata dal propranololo, un farmaco che blocca l'associazione tra adrenalina e recettore beta adrenergico. Il sistema dell'adenilciclasi è implicato in tutta una serie di

Le modalità d'azione di un mediatore sono esemplificate dall'effetto della noradrenalina su una ghiandola epifisaria. La noradrenalina (*punti rossi*) liberata dalle terminazioni della fibra nervosa si lega a un recettore beta adrenergico situato sulla superficie della cellula epifisaria. Il recettore, di conseguenza, attiva l'enzima adenilciclasi sul lato interno della membrana della cellula epifisaria. L'adenilciclasi attivata catalizza la trasformazione dell'adenisintrifosfato (ATP) nell'adenosinmonofosfato ciclico (AMP). L'AMP ciclico stimola la sintesi dell'enzima N-acetiltransferasi; l'enzima trasforma la serotonina in N-acetilserotonina. Questa è, a sua volta, trasformata nell'enzima specifico della cellula epifisaria, idrossindolo-O-metiltransferasi (HIOMT) che sintetizza melatonina, l'ormone della ghiandola epifisaria che agisce in maniera non del tutto chiarita sulle ghiandole sessuali.

azioni biologiche. L'epifisi compie la sua funzione specifica, l'elaborazione della melatonina, utilizzando il sistema quasi universale dell'adenilciclasi, solo se sulla superficie cellulare vi sono recettori che possono riconoscere in modo specifico la noradrenalina e se l'enzima idrossindolo-O-metiltransferasi, presente soltanto nelle cellule epifisarie, trasforma l'N-acetilserotonina in melatonina.

Dopo che il neuromediatore ha interagito con la cellula postsinaptica, la sua azione deve essere rapidamente bloccata altrimenti eserciterebbe il suo effetto troppo a lungo e comprometterebbe la possibilità di un preciso controllo. Nel sistema nervoso colinergico l'acetilcolina viene rapidamente inattivata dall'enzima acetilcolinesterasi, che metabolizza il mediatore. Negli ultimi dieci anni si è potuto tuttavia stabilire che l'inattivazione dei neuromediatori attraverso la trasformazione enzimatica rappresenta l'eccezione piuttosto che la regola. I neuromediatori catecolamminici sono metabolizzati da due enzimi, la catecolo-O-metiltransferasi (COMT) e la monoamminossidasi (MAO); quest'ultima è un enzima di particolare importanza che sposta il gruppo amminico (NH_2) da un'ampia serie di composti tra cui: serotonina, noradrenalina, dopammina e adrenalina. Esistono sostanze chimiche, dotate di azione inibente l'attività degli enzimi, che possono bloccare le trasformazioni biochimiche controllate da COMT e MAO: tuttavia, quando vengono somministrati questi inibitori, l'azione della noradrenalina continua a essere rapidamente inattivata. Deve quindi esistere un altro metodo di inattivazione rapida, oltre a quello delle trasformazioni enzimatiche.

Per determinare tale meccanismo l'autore ha iniettato noradrenalina radioattiva in un gatto. Il mediatore marcato persisteva per molte ore nei tessuti riccamente innervati da fibre simpatiche, anche molto tempo dopo che era cessata l'azione fisiologica, dimostrando che la noradrenalina radioattiva era assunta e conservata nei nervi simpatici. L'autore e i suoi collaboratori hanno effettuato un semplice esperimento per dimostrare questo fenomeno: in ratti si sono distrutti i nervi simpatici che innervano la ghiandola salivare sinistra, mediante l'asportazione del ganglio cervicale superiore, situato sul lato sinistro del collo. Sette giorni dopo questo intervento, mentre i nervi ricchi di noradrenalina della ghiandola salivare erano intatti sul lato destro, risultavano distrutti su quello sinistro. Iniettando noradrenalina radioattiva potevamo stabilire che il mediatore era presente nella ghiandola salivare destra, ma non in quella di sinistra. Si è anche osservato che nei gatti a cui era iniettata noradrenalina radioattiva, il mediatore veniva liberato quando si stimolavano elettricamente i nervi simpatici. In base a questi risultati si è ipotizzato che la noradrenalina venga rapidamente inattivata attraverso un suo recupero da parte del sistema simpatico; quando il neuromediatore ritorna nelle fibre nervose non può, ovviamente, esplicare alcun effetto sulle cellule postsinaptiche. Leslie L. Iversen, dell'Università di Cambridge, ha poi potuto stabilire che questo « recupero » da parte delle fibre nervose simpatiche è altamente selettivo per la noradrenalina o per i composti che presentano una struttura chimica similare. Lavori compiuti di recente indicano che l'assunzione da parte delle fibre nervose può rappresentare il meccanismo più generale di inattivazione dei neuromediatori.

Regolazione

I mediatori chimici nelle fibre simpatiche (e probabilmente anche nelle altre) si trovano in uno stato dinami-

co: sono continuamente sintetizzati, liberati, metabolizzati e recuperati. L'attività dei nervi può anche presentare marcate variazioni durante periodi di stress. Tuttavia, malgrado tutte queste modificazioni dinamiche, il contenuto tessutale in catecolammine rimane costante. Ciò è dovuto a numerosi meccanismi adattativi che modificano la formazione, liberazione e reazione delle cotecolammine.

Quando si stimolano i nervi simpatici aumenta rapidamente in essi la trasformazione di tirosina in noradrenalina. L'accresciuta attività nervosa interessa in maniera specifica la tirosinaidrossilasi, l'enzima che trasforma la tirosina in DOPA, perché la sua attività è inibita dalla noradrenalina e dalla dopammina. Ciascun incremento della stimolazione nervosa indotta da stress, freddo e farmaci riduce il livello delle catecolammine nelle terminazioni nervose. Ciò diminuisce l'effetto del meccanismo a retroazione negativo di noradrenalina e dopammina sulla tirosinaidrossilasi, in modo che una quantità maggiore di tirosina è trasformata in DOPA, che a sua volta è convertita per elaborare più catecolammine. Reciprocamente, quando l'attività nervosa diminuisce, il livello delle catecolammine aumenta, rallentando la trasformazione della tirosina in DOPA e inibendo ancora una volta la tirosinaidrossilasi. Un'altra regolazione rapida è effettuata a livello delle terminazioni nervose stesse, dove sono localizzati i recettori alfa adrenergici. Quando questi recettori sono attivati, riducono la liberazione di noradrenalina dalle terminazioni nervose alla fessura sinaptica. Se viene liberata una quantità eccessiva di noradrenalina, essa si accumula nella fessura sinaptica; qualora il livello delle catecolammine sia sufficientemente elevato, attiva gli alfa recettori sulle terminazioni nervose presinaptiche e blocca l'ulteriore liberazione di neuromediatore.

Un processo di regolazione più lento è innescato da una stimolazione prolungata dei nervi simpatici, che può accrescere la formazione di enzimi capaci di sintetizzare le catecolammine: tirosinaidrossilasi, DBH e, seppure in minor grado, PNMT; l'incremento del livello degli enzimi consente alle fibre nervose di elaborare una quantità maggiore di neuromediatori. L'autore ha osservato questo fenomeno di accresciuta sintesi enzimatica somministrando agli animali reserpina, una sostanza dotata di ampio spettro d'azione, che riduce la pressione sanguigna e, incidentalmente, aumenta la stimolazione dei nervi simpatici (che tende ad aumentare la pressione) attraverso una azione riflessa. La reserpina ha indotto un graduale incremento della tirosinaidrossilasi e della DBH nei nervi simpatici e nelle capsule surrenali e della PNMT nelle capsule surrenali. Un aumento di questi enzimi si è anche registrato negli animali esposti al freddo, a stress, tenuti in condizioni di limitata attività fisica, sottoposti a stimoli psicosociali o a cui è stata iniettata insulina. D'altro lato, se la sintesi proteica viene inibita da farmaci, non si registra alcun incremento dell'attività enzimatica dopo somministrazione di reserpina. Ciò dimostra che un'accresciuta attività nervosa stimola la sintesi di nuove molecole di tirosinaidrossilasi, DBH e PNMT: a un maggior fabbisogno di neuromediatori corrisponde un incremento compensatorio nella sintesi degli enzimi che catalizzano la formazione dei mediatori stessi.

Per chiarire se lo stimolo ad aumentare la sintesi di nuove molecole di tirosinaidrossilasi e DBH poteva essere trasmesso da una fibra all'altra, l'autore ha sezionato le fibre che innervano alcuni corpi cellulari noradrenalinici – i gangli cervicali superiori – di un solo lato. Quando poi le fibre nervose venivano stimolate per via riflessa dalla reserpina, si registrava un aumento del livello della tirosinaidrossilasi e della DBH nei gangli innervati, ma non in quelli denervati. L'esperimento ha dimostrato che una sola fibra nervosa può trasmettere informazioni a un'altra (probabilmente mediante uno stimolo chimico); ciò provoca la formazione di nuove molecole enzimatiche nella fibra postsinaptica.

Liberazione e inattivazione del neuromediatore noradrenalina. L'enzima DBH è immagazzinato nelle terminazioni nervose noradrenergiche insieme al mediatore ed è liberato con esso nella fessura sinaptica. La noradrenalina si lega ai recettori sulla cellula effettrice, provocandone la reazione come mostrato nella pagina precedente. In seguito l'azione della noradrenalina è inattivata, sia per via metabolica (*1*) dagli enzimi catecolo-O-metiltransferasi (COMT) e/o dalla monoamminossidasi (MAO), sia per recupero e immagazzinamento (*2*) nella terminazione presinaptica; quest'ultimo rappresenta il processo più importante. La MAO, immagazzinata nella membrana dei mitocondri, è pure in grado di inattivare la noradrenalina delle vescicole (*3*).

Sensibilità

Nel 1855 il fisiologo tedesco J.L. Burge osservò che, distruggendo nei conigli le fibre che innervano l'occhio destro, la pupilla di quest'occhio risultava più dilatata di quella controlaterale. Questo fenomeno fu più tardi chiarito dal fisiologo statunitense Walter B. Cannon secondo cui la denervazione rendeva più reattive le cellule effettrici. Egli definì questo effetto « ipersensibilità da denervazione ». Le ricerche successive spiegarono che questo fenomeno è causato da due meccanismi separati, l'uno pre- e l'altro post-sinaptico. Quando le fibre nervose sono distrutte, si abolisce l'inattivazione presinaptica del tipo a recupero, di conseguenza i neuromediatori reagiscono più a lungo con il sito postsinaptico. La denervazione provoca anche profonde variazioni nel grado di attività della cellula postsinaptica. Lavori recentemente compiuti nel laboratorio dell'autore sulla ghiandola epifisaria hanno permesso di formulare ipotesi che chiariscono la iper- e anche la iposensibilità nelle cellule postsinaptiche. Come è stato osservato, la noradrenalina stimola la sintesi dell'enzima serotonina N-acetiltransferasi attraverso un recettore beta adrenergico nella cellula epifisaria postsinaptica. Quando le fibre nervose afferenti delle cellule epifisarie sono distrutte (o impoverite del contenuto di noradrenalina tramite la reserpina), le cellule dell'epifisi aumentano di dieci volte la loro capacità di reagire alla noradrenalina; cioè, quando la cellula postsinaptica è privata del suo neuromediatore per un certo periodo di tempo, richiede un solo pacchetto di catecolammine per produrre lo stesso aumento di N-acetiltranferasi che produrrebbero dieci pacchetti di mediatori in una cellula normalmente innervata. Viceversa, se la cellula epifisaria è esposta a una quantità eccessiva di catecolammine per un certo periodo di tempo, diviene meno reattiva e richiede una quantità più elevata di mediatore.

Questi esperimenti dimostrano che le variazioni nella capacità di reagire delle cellule eccitabili sono la conseguenza di un'alterazione nel grado di « avidità » del recettore verso i neuromediatori. Se il recettore è esposto a piccoli quantitativi di catecolammine per un certo periodo di tempo, reagisce rapidamente con il neuromediatore mentre diviene meno reattivo se viene bombardato da un numero eccessivo di sue molecole. In relazione al tessuto, questa variazione di sensibilità può verificarsi entro ore o giorni, in modo da rappresentare un efficace meccanismo di adattamento per le cellule capaci di eccitarsi. È possibile che l'assuefazione spesso sviluppata verso un farmaco assunto in dosi eccessive possa riflettere uno stato di iposensibilità nei recettori cellulari che reagiscono a esso.

Neuromediatori cerebrali

Nei miliardi di neuroni cerebrali agiscono numerosi neuromediatori dotati di differenti caratteristiche, ma poiché si dispone di metodi più precisi per determinare le catecolammine e i farmaci capaci di alterarne la formazione, l'immagazzinamento, la liberazione e il metabolismo, questi neuromediatori sono più conosciuti di altri. La microfotografia a fluorescenza e i farmaci che distruggono in maniera selettiva le fibre nervose contenenti catecolammine hanno consentito di localizzare i corpi cellulari contenenti noradrenalina, dopammina e serotonina, tracciare il decorso dei loro assoni e definire le terminazioni nervose (*si veda l'illustrazione a pagina 51*). I corpi cellulari delle fibre contenenti dopammina sono localizzati nell'area del tronco cerebrale denominata sostanza nera da cui si dipartono gli assoni dopamminergici che attraversano il tronco cerebrale terminando, per la maggior parte, nel nucleo caudato. In questo nucleo i segmenti contenenti dopammina svolgono un'importante funzione nell'integrazione dell'attività motoria. La spiegazione dell'attività biochimica di questo neuromediatore nel cervello ha permesso di elaborare un trattamento molto efficace del morbo di Parkinson. Il farmacologo svedese Arvid Carlsson aveva notato, nel 1959, che, somministrando reserpina ai ratti, si riduceva in modo marcato il contenuto in dopammina del nucleo caudato e si aveva la comparsa di un tremore simile a quello osservabile nel morbo di Parkinson. La somministrazione di DOPA, un precursore della dopammina che può passare dal sangue al cervello più rapidamente di quest'ultima, eliminava i tremori. Queste osservazioni spinsero Oleh Hornykiewicz, che allora lavorava all'Università di Vienna, a determinare il contenuto in dopammina nel cervello di pazienti deceduti per morbo di Parkinson. Poté così rilevare che in questi casi non esisteva praticamente dopammina nel nucleo caudato. Sulla base di questi dati, George C. Cotzias, del Brookhaven National Laboratory, poté migliorare in modo radicale la terapia di questa forma morbosa: infatti somministrando DOPA, il precursore della dopammina che supera la barriera emato-encefalica, si elimina la deficienza in dopammina e si ha una effettiva remissione dei sintomi del morbo di Parkinson.

Nel cervello esistono due principali cordoni nervosi contenenti noradrenalina, il fascio dorsale e ventrale. I corpi cellulari dei tratti contenenti noradrenalina sono situati nella porzione cerebrale inferiore, nella zona definita *locus coeruleus*. Questi tratti sono molto ramificati e raggiungono numerose sezioni del cervello. Fra le aree innervate vi sono il cervelletto, la corteccia cerebrale, interessata alla fine coordinazione del movimento, alla veglia e all'emotività. Un'altra porzione del cervello innervata da neuroni contenenti noradrenalina è l'ipotalamo, che controlla numerosi funzioni corporee, come la fame, la sete, la regolazione della temperatura corporea, la pressione sanguigna, la riproduzione, il comportamento. L'alterazione dei livelli cerebrali della noradrenalina può modificare numerose attività funzionali ipotalamiche, in particolare i centri del « piacere ». I segmenti nervosi con noradrenalina sembrano implicati nell'esaltazione e depressione affettiva. Di recente sono state riscontrate fibre nervose nel tronco cerebrale contenenti adrenalina. I prossimi anni dimostreranno se questi tratti nervosi controllano anche l'emotività, l'affettività e il comportamento.

I farmaci si sono dimostrati un potente strumento di analisi dell'azione dei neuromediatori. L'approfondimento delle conoscenze relative a queste sostanze ha permesso di comprendere le modalità d'azione dei farmaci sul comportamento, sul sistema cardiovascolare e su quello motorio; le due linee di ricerche hanno sottilmente interagito. Agli inizi degli anni '50 i farmacologi hanno stabilito che un agente allucinogeno, la dietilammide dell'acido lisergico (LSD) non solo presenta analogie nella struttura chimica con la serotonina, ma ne contrasta talune azioni farmacologiche (occupando i siti stabiliti per la serotonina). Molti ricercatori hanno quindi dedotto che questa deve essere implicata in qualche modo con talune affezioni mentali. Anche altri agenti allucinogeni, come la mescalina e le anfetamine, sono strutturalmente correlati con la noradrenalina. Verso la metà degli anni '50 i ricercatori clinici incominciarono a constatare che talune sostanze chimi-

che, del tipo della cloropromazina, potevano migliorare un comportamento psicotico e che gli inibitori della monoamminossidasi, l'imiprammina e farmaci dello stesso tipo potevano attenuare gli stati depressivi. All'incirca nello stesso periodo si rilevò che la reserpina, dimostratasi valida non solo nella terapia dell'ipertensione, ma anche in quella della schizofrenia, riduceva in modo marcato i livelli cerebrali della noradrenalina e serotonina. Tutte le osservazioni convergevano nel dimostrare che questi farmaci esplicavano la loro azione sul cervello interferendo con i neuromediatori. Quando l'autore e i suoi colleghi riscontrarono che la noradrenalina poteva essere assunta e liberata dalle fibre nervose si trovarono in una posizione privilegiata per studiare come un farmaco potesse influenzare la distribuzione dei mediatori radioattivi iniettati.

Azione di droghe e farmaci

Il primo composto esaminato fu la cocaina, una potente sostanza stimolante che può produrre psicosi e che accentua l'azione della noradrenalina. Quando la noradrenalina radioattiva veniva somministrata in gatti cui era stata data cocaina, l'assunzione delle catecolammine da parte dei nervi simpatici risultava inibita, dimostrando in tal modo che la cocaina accentuava l'effetto della noradrenalina, bloccando il suo legame e l'inattivazione, e permettendo che quantitativi più elevati di catecolammina reagissero con le cellule effettrici. Il farmaci antidepressivi del tipo dell'imiprammina rivelano lo stesso tipo di comportamento: bloccano l'assunzione di noradrenalina nei nervi simpatici. Utilizzando noradrenalina radioattiva si poté constatare che l'anfetamina, la quale stimola e altera le attività mentali, interessa le fibre noradrenergiche in due modi: blocca l'assunzione di noradrenalina e favorisce la liberazione dei neuromediatori dalle fibre nervose.

Anche molti farmaci, efficaci nel trattamento dell'ipertensione, influenzano l'immagazzinamento e la liberazione dei mediatori noradrenergici. La reserpina e la guanetidina riducono la pressione sanguigna inibendo le fibre nervose che sono responsabili del suo aumento a causa dell'accumulo di noradrenalina. I farmaci antiipertensivi, del tipo del metil-DOPA, d'altra parte, sono trasformati per via enzimatica entro le fibre nervose in sostanze chimicamente somiglianti alla noradrenalina. I falsi mediatori sono immagazzinati e liberati insieme ai neuromediatori naturali, diluendoli e riducendone in tal modo l'effetto.

Taluni farmaci o droghe adrenergiche aumentano o diminuiscono la disponibilità della noradrenalina ai recettori adrenergici. Vengono illustrati la liberazione normale, il recupero e il metabolismo (*frecce colorate*), con un andamento che rappresenta la reazione normale di una cellula postsinaptica (*a*). I farmaci antidepressivi aumentano tale reazione in parecchi modi, tutti diretti ad accrescere la disponibilità di noradrenalina nella sinapsi. L'anfetamina agisce provocando la liberazione di noradrenalina (*b*). L'anfetamina, l'imiprammina e i composti correlati bloccano il recupero (*c*); gli inibitori della monoamminossidasi interferiscono con il processo di inattivazione attraverso il metabolismo (*d*). Viceversa, la reserpina che riduce la pressione sanguigna, diminuisce la reazione esaurendo la noradrenalina accumulata (*e*); l'alfametildopa e gli altri «falsi mediatori» vengono immagazzinati nelle vescicole con la noradrenalina insieme alla quale sono liberati diminuendone l'effetto (*f*).

Negli ultimi dieci anni molti psichiatri e fisiologi sono stati colpiti dall'osservazione che i farmaci capaci di migliorare gli stati depressivi interferiscono con l'assunzione, la liberazione e il metabolismo della noradrenalina. Mentre l'imiprammina blocca l'assunzione della noradrenalina da parte delle fibre nervose e l'anfetamina libera noradrenalina e ne blocca l'assunzione, gli inibitori della monoamminossidasi, appunto come dice il loro nome, inibiscono il metabolismo delle catecolammine. In altre parole, tutti questi antidepressivi provocano risultati simili attraverso meccanismi diversi: aumentano il quantitativo delle catecolammine nella fessura sinaptica, con la conseguenza di creare una maggiore quantità di mediatore disponibile per stimolare il recettore. Viceversa, la reserpina, una sostanza che riduce la quantità dei mediatori chimici, talvolta produce forme depressive. Queste osservazioni hanno fatto formulare una ipotesi catecolamminica degli stati depressivi secondo cui la depressione mentale è associata a una diminuita disponibilità di catecolammine cerebrali ed è migliorata da farmaci che accrescono la quantità di questi mediatori a livello del recettore adrenergico. Anche se tale ipotesi non è stata ancora completamente dimostrata in sede sperimentale, ha tuttavia fornito una solida base su cui svolgere nuove ricerche per comprendere i meccanismi responsabili dell'insorgenza degli stati depressivi.

I principali fasci di nervi contenenti dopammina sono illustrati (*in nero*) in un disegno di una sezione longitudinale compiuta lungo la linea mediana del cervello di ratto. I corpi cellulari sono concentrati nella sostanza nera e gli assoni si proiettano primariamente verso il nucleo caudato. Una deficienza di dopammina in questa regione causa le manifestazioni del morbo di Parkinson, che può essere trattato con DOPA.

I tratti noradrenergici si originano nel *locus coeruleus* e raggiungono numerosi centri cerebrali, compresi il cervelletto, la corteccia cerebrale e l'ipotalamo. I disegni sono stati eseguiti in base a schizzi di Urban Ungerstedt, del Royal Caroline Institute.

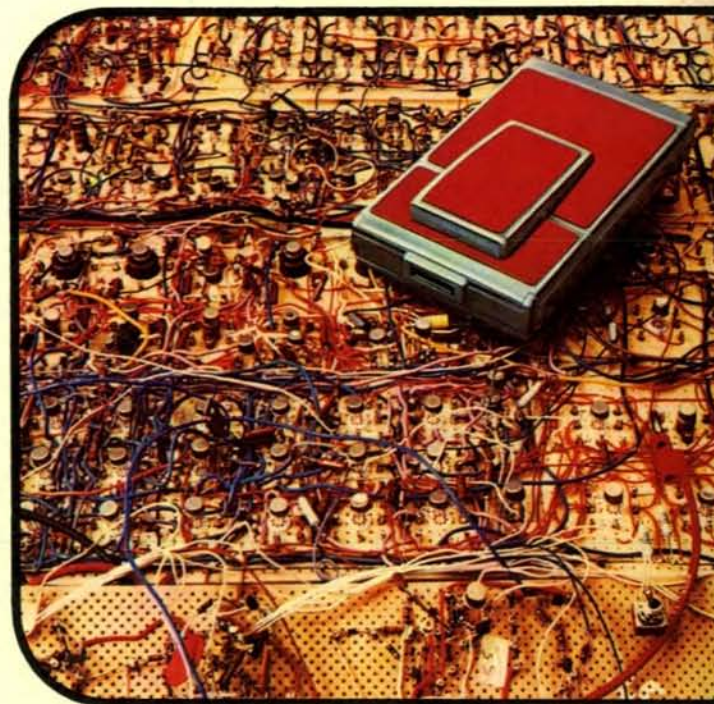
L'introduzione negli anni '50 di farmaci antipsicotici del tipo della cloropromazina e aloperidolo ha rivoluzionato il trattamento della schizofrenia, riducendo in modo rilevante la durata del periodo di ricovero e facendo economizzare miliardi in cure ospedaliere. Le ricerche svolte nel passato decennio hanno rivelato che anche tali farmaci svolgono la loro azione sui neuromediatori catecolamminici. Carlsson ha riscontrato che i preparati antischizofrenici inducono un'aumentata formazione di catecolammine nel cervello di ratto e formulò l'ipotesi che ciò si verifica per la capacità del farmaco di bloccare i recettori della dopammina. Anche le indagini svolte da altri ricercatori hanno confermato e ampliato questa ipotesi. I farmaci antipsicotici bloccano i recettori della dopammina nel cervello ed esiste un elevato grado di correlazione fra capacità di blocco dei vari farmaci e loro facoltà di migliorare i sintomi schizofrenici. Questi dati attestano in maniera evidente che nella schizofrenia sono interessate le fibre dopamminergiche. L'anfetamina ha anche consentito di chiarire la natura della schizofrenia. Somministrata ripetutamente, in dosaggi elevati, l'anfetamina produce una psicosi che si manifesta con un comportamento stereotipo e compulsivo e allucinazioni indistinguibili dai sintomi propri della schizofrenia paranoide. L'anfetamina libera le catecolammine dalle fibre nervose del cervello per stimolare sia i recettori adrenergici sia quelli dopamminergici. Dopo avere compiuto esperimenti con due forme di anfetamina, Solomon H. Snyder della Scuola di medicina dell'Università Johns Hopkins, formulò l'ipotesi che la psicosi schizofrenicasimile indotta dalla droga è dovuta a un'eccessiva liberazione di dopammina. La capacità dei farmaci antischizofrenici, che bloccano i recettori della dopammina, a mitigare la psicosi da anfetamina concorda con questa ipotesi. Per quanto negli ultimi vent'anni siano stati compiuti rapidi progressi nelle conoscenze relative ai neuromediatori, molto deve essere ancora indagato su questi composti. Se si potrà invertire l'attuale tendenza a ridurre i fondi destinati alle ricerche saranno presto possibili nuove, rilevanti, scoperte sui neuromediatori, molte delle quali capaci di contribuire direttamente al trattamento o cura di talune delle più tragiche sofferenze umane.

e regolazione della messa a fuoco
dall'infinito fino a 25 cm dal fronte dell'apparecchio.
Essendo reflex, comunque, la possibilità di usare
lenti addizionali permette di fotografare a distanze
estremamente ravvicinate. Interessante è la soluzione
adottata per l'otturatore. Si tratta in pratica di una
coppia di lamelle che scorrono orizzontalmente
una contro l'altra, lasciando scoperta al centro
un'apertura attraverso la quale passa la luce. Lo stesso
otturatore, o meglio l'apertura centrale, ha anche la
funzione di diaframma, regolato dal diametro che questa
apertura raggiunge. Le stesse lamelle dell'otturatore
portano due fessure, analoghe a quelle che si trovano
davanti all'obiettivo, in corrispondenza della fotocellula,
che regola in questo modo il movimento di chiusura
delle lamelle stesse. L'esposizione è automatica
da 1/180 di secondo a 14 secondi e da f/8 a f/96.
In casi particolari di ripresa (per esempio in controluce)
è possibile regolare l'esposizione su valori inferiori
o superiori a quelli dettati dalla fotocellula.
Anche usando il lampo (anch'esso di nuovo tipo,
a barretta, appositamente studiato per questo
apparecchio) l'esposizione è automatica: fisso il tempo
di posa (1/40 di secondo) l'apertura del diaframma
viene impostata regolando la ghiera di messa a fuoco.

COME FUNZIONA

Al momento dello scatto (anche questo
a comando elettrico e non meccanico) l'otturatore,
normalmente aperto per permettere la visione attraverso
l'obiettivo, si chiude. Si alza la piastra
che si trova alla base dell'apparecchio per proteggere
la pellicola dalla luce, si apre l'otturatore in modo da
permettere l'esposizione, la pellicola viene poi
agganciata ed espulsa dalla parte frontale
dell'apparecchio per mezzo di un motorino.
Infine si riabbassa la piastra e si riapre l'otturatore.
Al funzionamento dell'apparecchio provvedono circuiti
integrati che corrispondono a oltre 200 transistori,
e la fonte di energia (la pila) si trova
in ogni confezione di pellicola. Soluzione anche questa
estremamènte nuova, che mette al riparo da eventuali
pericoli di dimenticare di sostituire le pile scariche,
o di non trovarle al momento giusto. La pila, 6 V e 8 A,
è ultrapiatta e trova comodamente posto
alla base del filmpack che contiene le pellicole.

LA PELLICOLA

Le soluzioni tecnologicamente più avanzate
si hanno comunque nella pellicola.
Mentre per l'apparecchio si può parlare di
perfezionamenti di tecniche già note, per la pellicola si
tratta di una vera e propria rivoluzione: in meno di 2 mm
si trovano ben 17 strati di cui alcuni dello spessore di
qualche micron. Questo già comporta la soluzione
di problemi non indifferenti dal punto di vista della
produzione industriale, che deve garantire uno stardard
qualitativo estremamente elevato. Ma è il processo
negativo-positivo che rompe decisamente i ponti con la
fotografia tradizionale. Infatti, mentre si può
ancora parlare di una immagine negativa di tipo comune,
costituita da tre strati di alogenuri d'argento
sensibili ai tre colori primari (blu, verde e rosso),
la positiva si forma non più su altrettanti strati
sensibili positivi, ma per diffusione dei coloranti
complementari già contenuti nella negativa,
che migrano, dopo l'esposizione, verso uno strato
superiore destinato ad accoglierli. L'immagine,
di 8,2 ×8,2 cm è bordata. Alla base un bordo più largo
contiene i liquidi di sviluppo; i bordi più sottili, oltre che
svolgere una funzione decorativa, contengono eventuali
eccessi di sviluppo che potrebbero fuoriuscire.
Sotto un sottile foglio di materiale plastico trasparente,
che assicura tra l'altro una sufficiente rigidità
alla fotografia, si trova uno strato di polimero acido,
poi uno strato di polimero detto « a tempo »,
poi lo strato destinato a ricevere i coloranti positivi,
poi uno di gelatina. Sotto a questi la negativa,
costituita dai tre strati sensibili al blu, verde e rosso
e ai corrispondenti coloranti metallizzati
di colore complementare. Alla base, infine, uno strato
protettivo opaco alla luce. Dopo l'esposizione,
la fotografia viene espulsa dalla parte frontale
dell'apparecchio, passando attraverso due rulli d'acciaio
che provocano la stesa uniforme dei reagenti
nello strato destinato ad accoglierli sopra la negativa.
Questi reagenti, sostanzialmente identici
a quelli delle precedenti pellicole Polaroid,
contengono una sostanza fortemente alcalina, biossido
di titanio, e una sostanza che impedisce il passaggio
della luce alla negativa,
che altrimenti annerisce completamente, permettendo

lo svolgersi delle ulteriori fasi di sviluppo
anche in piena luce. Da questo momento
le molecole dei coloranti in corrispondenza degli
alogenuri d'argento impressionati dalla luce
in ciascuno dei tre strati migrano verso l'alto,
completando la formazione dell'immagine positiva
nel giro di pochi minuti.

GLI ACCESSORI

A completare la versatilità dell'apparecchio
esiste una serie di accessori
appositamente creati per la SX-70. Particolarmente utili
sono la staffa e il comando a distanza,
oltre alla lente addizionale per la macrofotografia.

Cicli di popolazione nei roditori

Le popolazioni di molti piccoli roditori aumentano e diminuiscono secondo un ciclo di tre, quattro anni. A queste fluttuazioni si associano alcune tenui variazioni periodiche del patrimonio genico

di Judith H. Myers e Charles J. Krebs

Vi è una leggenda secondo la quale i lemming della Scandinavia aumentano periodicamente di numero in maniera straordinaria, risolvendo il loro ploblema di sovrappopolazione con lunghe migrazioni che si concludono con un suicidio in massa nel mare. In realtà le popolazioni di lemming e di altri piccoli roditori come le arvicole e i topi campagnoli fluttuano di anno in anno e una popolazione che è cresciuta costantemente per alcune stagioni può poi declinare improvvisamente e quasi scomparire del tutto. Si è anche stabilito di recente che le migrazioni, pur non concludendosi definitivamente in mare, costituiscono una parte essenziale del processo che regola le suddette popolazioni.

In molte specie le fluttuazioni sono cicliche con un periodo, tra i massimi di popolazione, che varia dai tre ai quattro anni. Oscillazioni di notevole ampiezza sono comuni: un ettaro di terreno che ospita più di duecento roditori nel periodo di massimo popolamento può, in alcuni periodi, ospitarne solo due-cinque o addirittura nessuno. Questo caratteristico ritmo è stato descritto in numerose specie della fascia artica e temperata dell'America settentrionale, dell'Europa e dell'Asia (invece, non sono stati notati cicli corrispondenti nei roditori dei paesi tropicali o dell'emisfero meridionale). I cicli sono relativamente regolari se misurati nell'arco di decenni e inoltre, di solito, avvengono contemporaneamente anche in comunità assai distanti tra loro. Non è stata ancora trovata una soddisfacente spiegazione di questo fenomeno. Per esempio, non si sa perché la lunghezza del ciclo sia quasi invariabilmente di tre-quattro anni, o perché popolazioni, chiaramente isolate l'una dall'altra, fluttuino in maniera sincrona. Le ricerche finora compiute sono state indirizzate verso la spiegazione delle cause di queste fluttuazioni. Sta ora delineandosi un modello del meccanismo che mette in moto questi cicli.

Le prime indagini sistematiche di dinamica delle popolazioni nei roditori furono effettuate da Charles S. Elton dell'Università di Oxford. Egli fu uno dei primi a riconoscere la natura periodica delle fluttuazioni. Negli anni venti, quando Elton cominciò il suo lavoro, erano disponibili scarse prove dirette sui cicli di popolazione. Un'analisi della letteratura sull'argomento rivelò che la popolazione dei lemming comuni (*Lemmus lemmus*) nella Norvegia meridionale aveva avuto dei massimi nel 1862, 1866, 1868, 1871, 1875, 1879, 1883, 1887, 1890, 1894, 1897, 1902, 1906 e 1909. L'intervallo medio tra i massimi di popolazione era di 3,8 anni. I cicli erano tuttavia documentati soltanto dall'osservazione di orde di lemming, facilmente visibili in annate di massima densità.

Ulteriori prove si poterono ricavare indirettamente. Basandosi sul fatto che le dimensioni delle popolazioni dei roditori hanno un riflesso sul numero dei loro predatori, molti dei quali costituiscono voci importanti nel commercio delle pellicce, Elton prese in esame alcune statistiche compilate dalla Compagnia della Baia di Hudson e da altre compagnie che importavano pellicce dall'America settentrionale. Egli trovò che nel Labrador settentrionale il numero di pelli di volpe rossa, che si nutre prevalentemente di arvicole, varia con un periodo dai tre ai quattro anni; nell'area circostante la Baia di Hudson un ciclo analogo fu stabilito per la volpe polare, che preda soprattutto i lemming (*si veda l'illustrazione in alto a pagina 56*).

I dati di Elton erano, da un certo punto di vista, inesatti e incerti; per esempio, le statistiche del commercio delle pellicce potevano essere influenzate dalla moda, da considerazioni economiche e da altri fattori non correlati con l'abbondanza della preda. L'andamento dei cicli era tuttavia sufficientemente chiaro e regolare da permettere di concludere che, nei roditori, i cicli di popolazione sono un fenomeno diffuso, comune alle aree agricole dell'Europa e alle comunità artiche virtualmente indisturbate. Egli era particolarmente affascinato dal fatto che, nelle regioni artiche del Canada, molte popolazioni indipendenti, separate da notevoli distanze, fluttuavano in fase l'una con l'altra. Il ritmo pulsatorio di queste popolazioni era presumibilmente collegato con la disponibilità di cibo di origine vegetale di una determinata zona che, a sua volta, influiva sui roditori e sui loro predatori.

Verso la fine degli anni venti, Elton e i suoi collaboratori adottarono tecniche più raffinate per lo studio delle popolazioni dei roditori. Usando trappole ideate per catturare esemplari vivi, essi riuscirono a prendere lo stesso animale parecchie volte; marcando e mettendo il libertà gli animali intrappolati, furono in grado di seguire i cicli vitali di individui dalla chiara identità. A quell'epoca le malattie epidemiche erano considerate la causa principale dei cicli di popolazione ed Elton fece il tentativo di identificare la malattia specifica, responsabile dei cicli di popolazione del topo campagnolo britannico. Benché con i suoi collaboratori fosse riuscito a scoprire una nuova malattia in questo animale (la tubercolosi) dovette, tuttavia, respingere la suddetta ipotesi. In effetti furono rilevate alcune epidemie, ma esse non erano sempre cicliche: alcune popolazioni subirono un declino senza alcuna traccia di malattia.

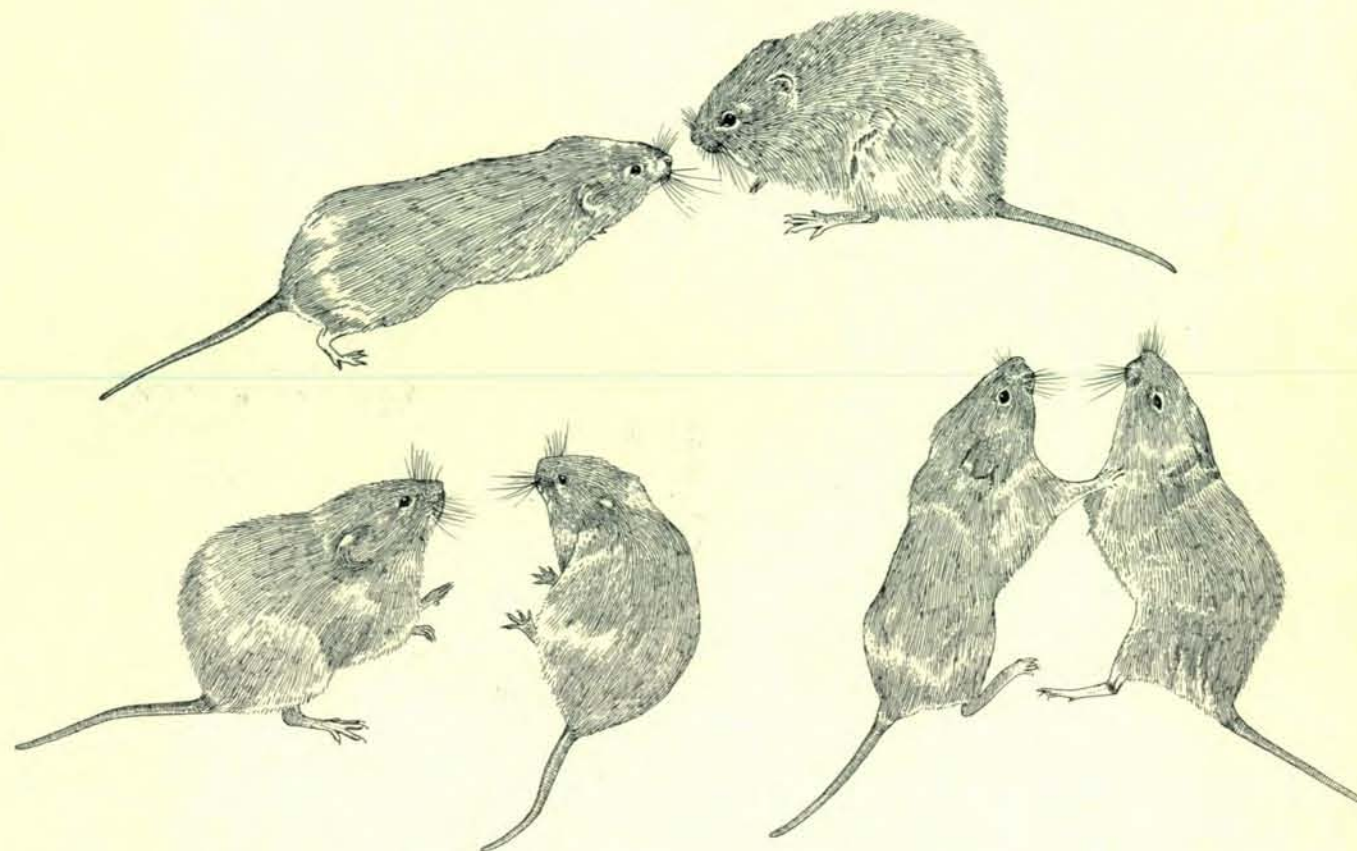
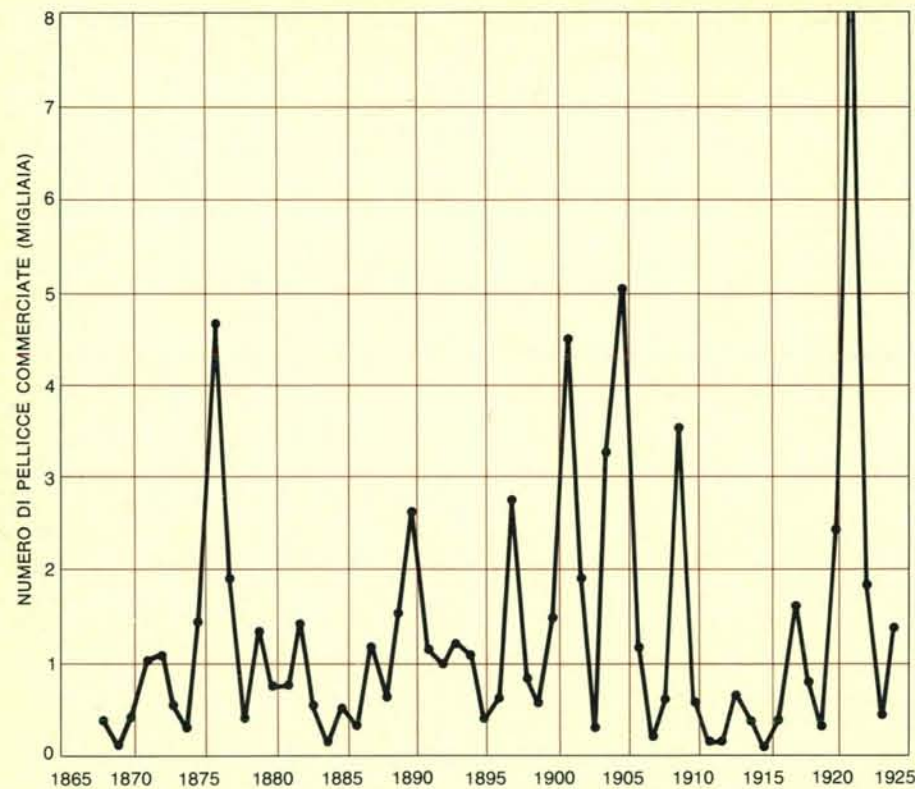
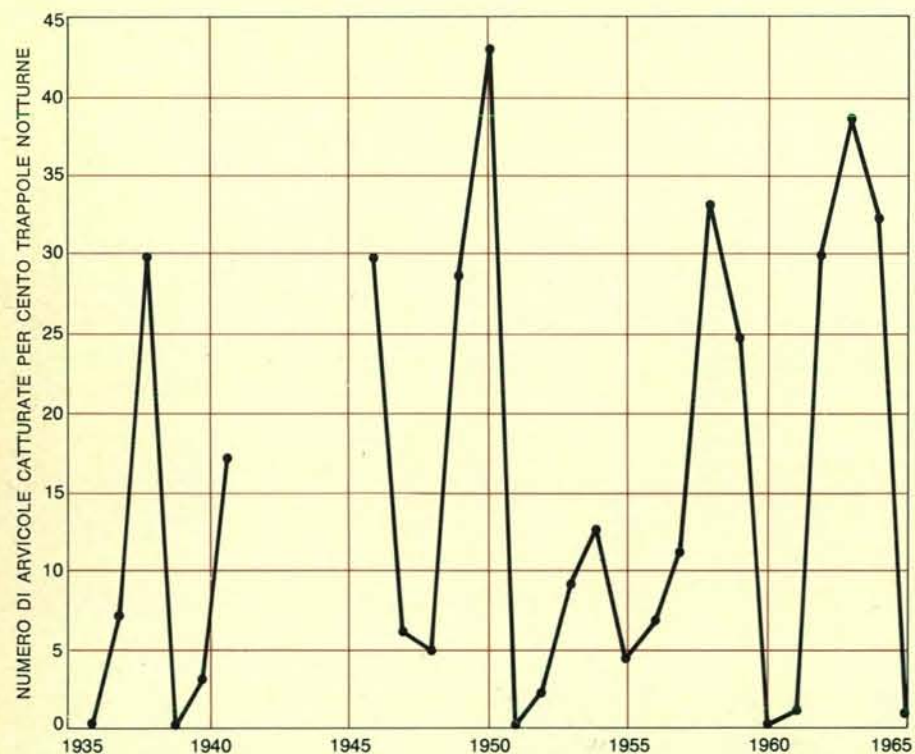
Il fallimento dell'ipotesi delle epidemie alla fine degli anni trenta lasciò gli ecologi senza una spiegazione soddisfacente per i cicli di popolazione dei roditori. Le due alternative più ovvie, predazione e carestia, furono esse pure contraddette dall'evidenza. Non sembrava che i predatori fossero abbastanza numerosi da determinare l'eliminazione delle numerose popolazioni che costituivano le punte massime dei cicli e, d'altra parte, i roditori catturati nei periodi di declino non mostravano alcun segno evidente di denutrizione.

I primi dati consultati da Elton, per esempio le statistiche sul commercio delle pellicce, erano essenzialmente valutazioni annue. Campionando le popolazioni con maggior frequenza, abbiamo potuto descrivere cicli di popolazione nei roditori con un dettaglio molto superiore. Con questa tecnica è possibile non solo identificare gli anni di massima e di minima di popolazione, ma anche determinare, settimana per settimana, la velocità di aumento o di declino della popolazione per identificare tutte le fluttuazioni a breve termine che si sovrappongono al ciclo e poter stabilire in maniera categorica la composizione della popolazione in qualsiasi momento del ciclo.

Tutte le specie di roditori interessate in questi studi sono erbivore e si nutrono di una grande varietà di piante verdi. Esse hanno vita breve: pochi sono gli individui che sopravvivono più di un anno. Tuttavia, la loro velocità di riproduzione è molto elevata. Arvicole e lemming sono attivi tutto l'anno e d'inverno scavano cunicoli sotto la neve; molte specie si possono riprodurre in ogni stagione.

Un ciclo di popolazione tipico è quello di *Microtus pennsylvanicus* degli Stati Uniti centrali e orientali, che abbiamo studiato nei campi circostanti l'Indiana University (*si veda la figura in alto a pagina 57*). Tale ciclo incomincia nella tarda primavera o all'inizio dell'estate, in un periodo di popolamento minimo. In questo periodo, la densità è di 2-3 animali per ettaro; un aumento estremamente rapido porta la popolazione a una densità altissima nella primavera successiva. Alle volte il numero dei roditori può crescere del 10-15 per cento ogni settimana ed è proprio durante questo periodo di forte aumento della popolazione che la riproduzione può proseguire anche durante l'inverno.

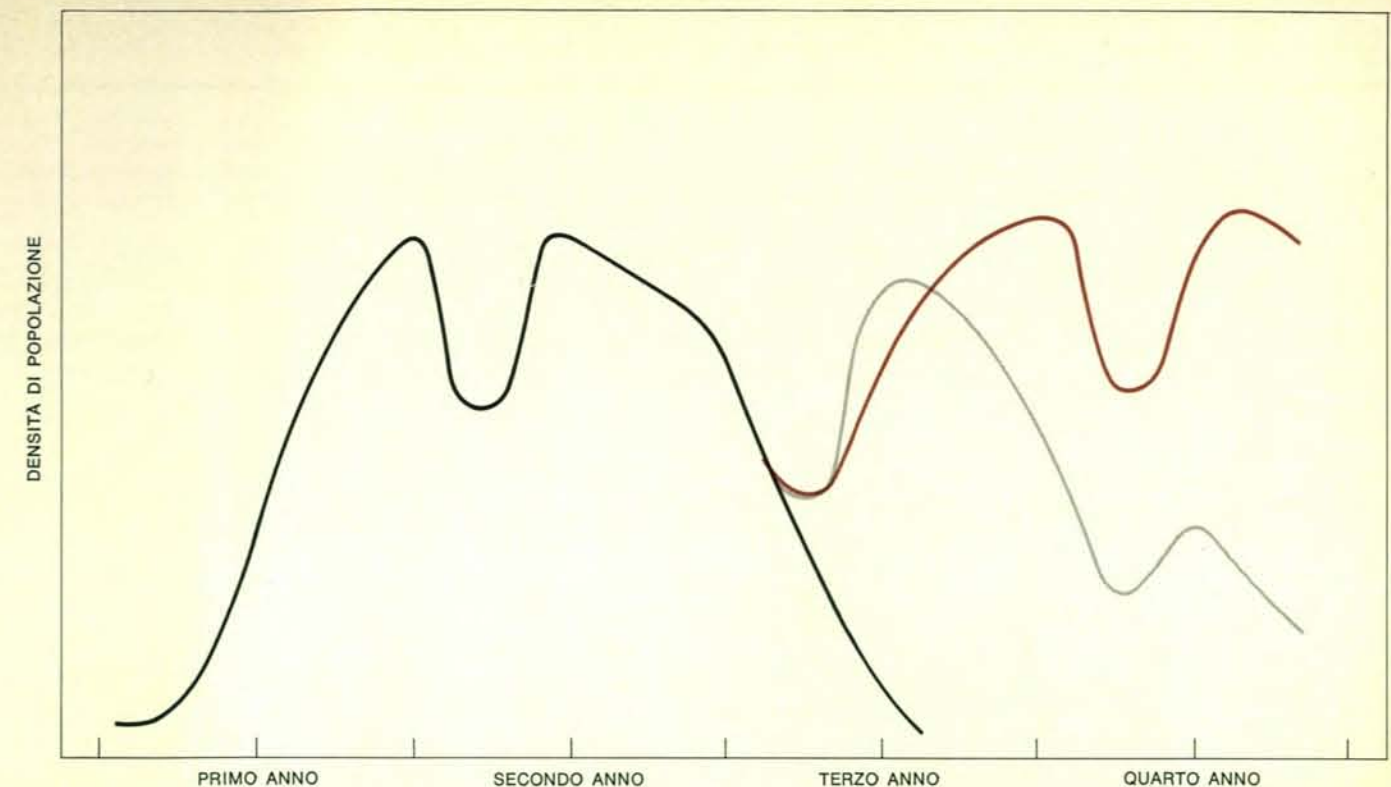
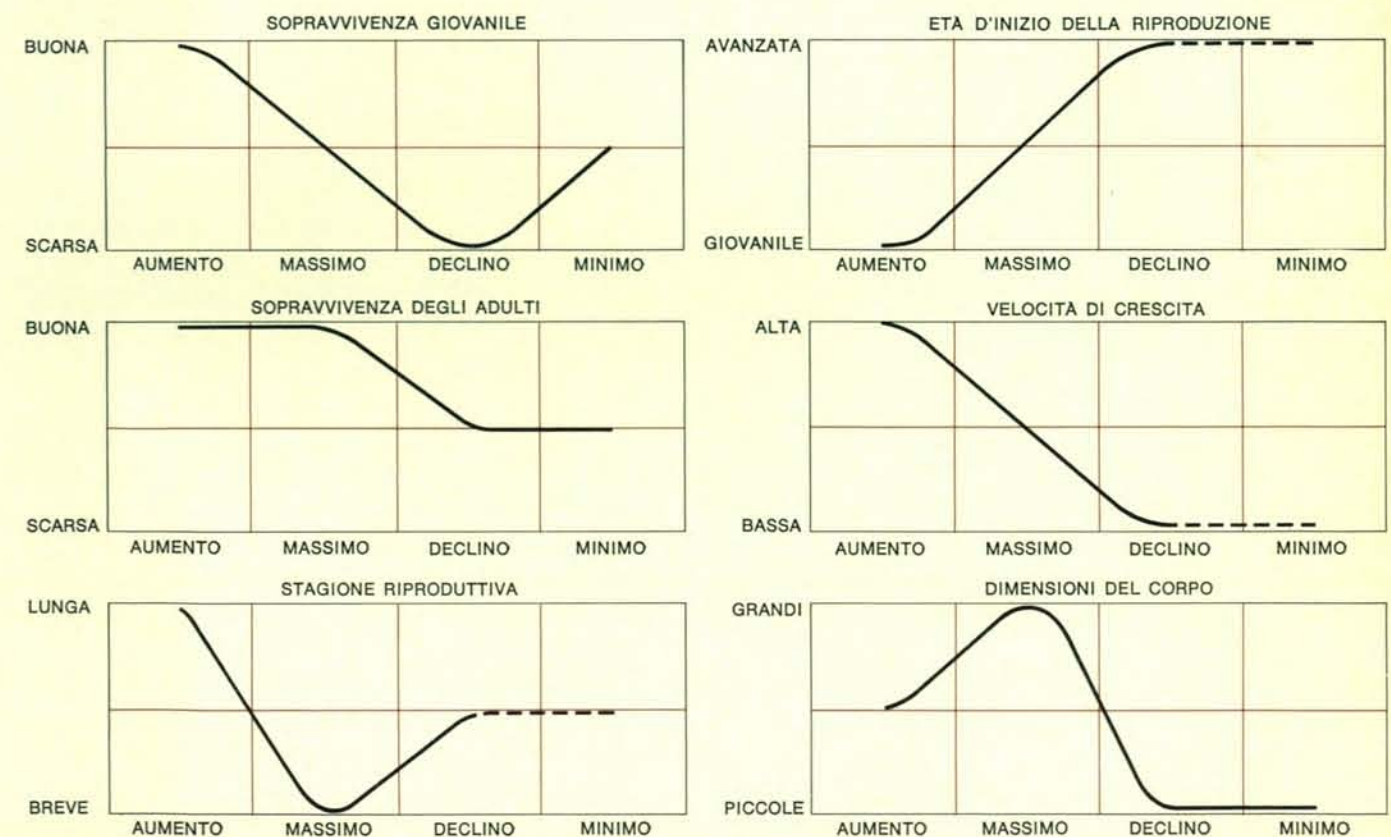
La densità massima raggiunta dalla comunità all'inizio della primavera dipende dalle condizioni locali. Può variare da più di 100 a più di 700 animali per ettaro, ma la cifra più frequente è di circa 250 (secondo gli standard di laboratorio, questa densità massima è piuttosto bassa: 250 animali per ettaro corrispondono a 2-3 animali in una stanza di 6 metri quadrati). In primavera si verifica spesso un'improvvisa diminuzione della densità e la popolazione può anche dimezzarsi; questa mortalità è di solito più elevata tra i maschi. In estate la densità della popolazione comincia ad aumentare e in autunno ritorna generalmente a un livello prossimo al massimo primaverile.

La riproduzione rallenta durante l'autunno, per cessare completamente e non riprendere fino alla primavera successiva. La popolazione diminuisce in

L'aggressività delle arvicole è stata misurata nel laboratorio degli autori confinando due maschi in una piccola gabbia e osservando le loro reazioni. Il comportamento considerato aggressivo consisteva di una sequenza di atti definiti « approccio » (*in alto*), « minaccia » (*a sinistra*) e « attacco » (*a destra*). Ogni topo venne confrontato con almeno due contendenti. L'incidenza del comportamento aggressivo variava nel corso del ciclo di popolazione; un animale proveniente da una popolazione numerosa era più aggressivo di uno che proveniva da una popolazione scarsa. Le fluttuazioni nell'aggressività dimostrano che la selezione naturale agisce su caratteristiche determinate geneticamente che si esprimono in interazioni sociali.

Una prova indiretta di ciclo di popolazione nei roditori fu ricavata dai dati registrati dalla Compagnia della Baia di Hudson per il commercio delle pellicce nel Labrador settentrionale. I dati si riferiscono alla volpe polare, che si nutre primariamente di lemming, e mostrano un ciclo con un periodo di tre-quattro anni nel numero di pellicce commerciate. Le statistiche furono compilate negli anni venti da Charles S. Elton dell'Università di Oxford. Benché fattori diversi dall'abbondanza della preda potessero influenzare l'entità del commercio delle pellicce, la regolarità delle fluttuazioni suggerì a Elton l'idea di una variazione ciclica nelle popolazioni di roditori. Il numero di pellicce commerciate nel 1921 fu molto elevato raggiungendo i 9797 capi.

Rilievi annuali della popolazione dell'arvicola della penisola di Kola nell'URSS settentrionale rivelano un ciclo di 3-4 anni. Il metodo non dà la popolazione totale, ma la densità relativa. I dati vennero raccolti da Koshkina nella riserva di Kandalaksha.

inverno, ma solo di poco. Muoiono in genere solo gli individui più vecchi o quelli più deboli. Nella primavera successiva la popolazione cala rapidamente e può raggiungere un basso valore di densità all'inizio dell'estate. Il ciclo è così completo, ma ciò che accade in seguito è variabile e impossibile da prevedere. La popolazione può ricominciare immediatamente a crescere, oppure può continuare a subire un declino diventando così esigua che è difficile catturare anche un solo animale, oppure può rimanere a un livello relativamente basso senza aumentare o diminuire per un anno o due.

La mutevole composizione della popolazione secondo il sesso e l'età, altre caratteristiche osservate durante il ciclo, è stata descritta nei particolari per numerose specie (*si veda l'illustrazione in basso nella pagina a fronte*). La sopravvivenza sia degli adulti, sia dei giovani è buona durante il periodo di aumento della popolazione, ma la mortalità giovanile aumenta nel periodo di massima densità e di declino. La sopravvivenza degli adulti rimane invece elevata durante il periodo di massima densità mentre crolla rapidamente nel periodo di declino. La velocità di riproduzione varia costantemente da un massimo che si verifica durante l'aumento a un minimo durante il declino. La velocità di crescita dei singoli animali varia pure da un massimo durante la fase di aumento della popolazione a un minimo durante la fase di declino, con il risultato che le popolazioni a densità più elevate sono costituite da animali di dimensioni superiori alla media.

La scoperta di andamenti complessi nella fluttuazione dei tassi di natalità, di mortalità e di crescita in cicli di popolazione ha reso poco attendibili le spiegazioni tradizionali su questi cicli. Le malattie, la predazione o le carestie potrebbero spiegare alcuni di questi fenomeni, ma non tutti. Sembra che sia invece di grande importanza un meccanismo di controllo interno che regola le interazioni tra gli animali che compongono la popolazione.

Un nuovo modo di affrontare lo studio dei cicli di popolazione fu proposto nel 1950 da John J. Christian, che lavora attualmente all'Università di stato di New York a Binghamton. Hans Selye dell'Università di Montreal aveva dimostrato in precedenza che lo stress prodotto da agenti fisici e chimici può portare a disturbi del sistema endocrino in animali di laboratorio. La condizione che ne deriva è

Il ciclo di popolazione del topo campagnolo *Microtus pennsylvanicus* ha inizio con un periodo di crescita estremamente rapido durante il quale la riproduzione può prolungarsi spesso anche nel corso dell'inverno. Nel secondo anno la popolazione diminuisce improvvisamente in primavera, ma aumenta di nuovo in autunno, quindi declina gradatamente con il cessare della riproduzione. Nella primavera successiva, quando la riproduzione riprende, il declino diventa precipitoso. Può continuare fino a raggiungere una densità molto bassa (*in nero*); può essere graduale con parziale recupero in autunno, raggiungendo la densità minima il quarto anno (*in grigio*), oppure può arrestarsi riportando la popolazione ai valori originari (*in colore*).

La costituzione variabile della popolazione durante un singolo ciclo può essere messa in evidenza misurando la sopravvivenza degli adulti e dei giovani, la lunghezza della stagione riproduttiva, l'età della prima riproduzione, la velocità di crescita e le dimensioni degli individui. Le dimensioni delle nidiate e la percentuale di adulti riproduttori variano, ma non ciclicamente.

nota come malattia da shock e può concludersi con la morte. Christian aveva avanzato l'ipotesi che elevate densità di popolazione possono produrre uno stress e far accelerare il declino in animali a seguito di turbe a carico del sistema endocrino.

Alcuni effetti dello stress si notano, di fatto, nelle popolazioni di piccoli roditori che vivono nei campi, ciononostante vi era un grave errore di fondo nell'argomento sostenuto da Christian. Dennis Chitty, che lavorava nel gruppo di Elton a Oxford, mise in evidenza che gli effetti dell'alta densità si riscontravano non nella generazione di animali esposti all'affollamento, ma nella loro prole. Era ovviamente implicato un meccanismo ereditario e Chitty cercava di capire se lo stress associato ad altre condizioni nelle popolazioni con densità massima potesse essere selettivo dal punto di vista genetico. Egli avanzava l'ipotesi che le interazioni sociali in una comunità affollata potessero favorire la riproduzione da parte di individui che possedevano determinate caratteristiche genetiche e scoraggiare la riproduzione di altri. La modificazione della composizione genetica della popolazione altererebbe a sua volta il carattere delle interazioni sociali all'interno della comunità.

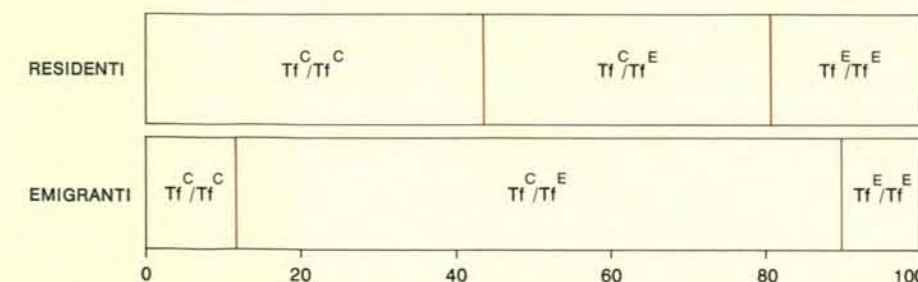
All'inizio l'ipotesi di Chitty fu accolta con scetticismo. Si riteneva che la selezione naturale operasse davvero troppo lentamente per poter influire su un ciclo con un periodo di soli pochi anni, ma le precedenti spiegazioni erano chiaramente inadeguate.

Nel 1965, cominciammo una serie di esperimenti intesi a mettere alla prova alcune delle previsioni implicite nell'ipotesi di Chitty. Una di queste era che le interazioni sociali all'interno della popolazione sarebbero cambiate durante il ciclo. In particolare decidemmo di indagare sulle modificazioni a livello dell'ostilità o aggressività tra individui.

L'aggressività fu valutata confinando due maschi in un piccolo recinto e notando le loro reazioni. Il repertorio del comportamento definito aggressivo consisteva in una serie di atti che abbiamo denominati avvicinamento, minaccia e attacco. La sottomissione a uno qualsiasi di questi atti, o la rinuncia da parte di ambedue gli animali ad avvicinarsi, a minacciare o ad attaccare, fu considerata come una prova di mancanza di aggressività. Ogni maschio fu messo a confronto con almeno due antagonisti così da permetterci di misurare tutta la gamma delle sue reazioni (*si veda la figura a pagina 55*).

Tra il 1965 e il 1970 studiammo 1140 maschi di *Microtus pennsylvanicus* e 1450 maschi di *Microtus ochrogaster*. I maschi delle popolazioni a elevata densità erano più aggressivi dei maschi di popolazioni in aumento o in declino. I cambiamenti di comportamento in risposta a una variazione della densità della popolazione confermarono le previsioni di Christian e di Chitty.

Per verificare l'ipotesi di Chitty era fondamentale determinare se la selezione naturale poteva operare abbastanza rapidamente da giustificare le variazioni osservate nel comportamento sociale. Fu quindi necessario misurare le variazioni genetiche nella popolazione. Dove i test di aggressività avevano riguardato un carattere fenotipico dei singoli individui, era ora necessario considerare il genotipo. Quando si ha a che fare con caratteri complessi trasmessi per via genetica, come i tipi di comportamento animale, può essere difficile eseguire una determinazione del genere, poiché il carattere è controllato non da un singolo gene (come è, per esempio, il colore dei semi nei piselli), ma da parecchi o molti geni che si possono esprimere in un individuo in varie combinazioni. Pertanto non è possibile esaminare semplicemente l'animale per identificarne il genotipo.

Per studi come il nostro è indispensabile avere a disposizione un indicatore sensibile alle differenze nel genotipo: con la tecnica dell'elettroforesi è possibile individuare nella composizione chimica delle proteine del sangue lievi differenze. La composizione in amminoacidi di certe proteine varia lievemente da individuo a individuo; queste variazioni hanno presumibilmente uno scarso effetto sul funzionamento delle proteine, ma possono alterare l'equilibrio delle cariche elettriche sulle loro molecole. L'elettroforesi sfrutta tali piccole differenze di carica. Le proteine ottenute dal siero di sangue sono poste su gel su cui si crea una differenza di potenziale elettrico. Le molecole dotate di una carica più intensa migrano più rapidamente attraverso il gel e formano bande separate da quelle delle molecole che migrano più lentamente. Dato che le proteine rappresentano un'espressione diretta di informazione genetica, le bande che si formano per effetto dell'elettroforesi possono servire come indicatori attendibili dei genotipi. È importante notare che queste lievi variazioni nella

L'elettroforesi discrimina tra diversi genotipi mettendo in luce lievi differenze nella composizione, specificata geneticamente, delle molecole proteiche. In questi studi sulla leucina-amminopeptidasi (*a sinistra*) e sulla transferrina (*a destra*) di arvicola, campioni di siero di sangue sono stati posti a un'estremità di un blocco di gel. Creando una differenza di potenziale attraverso il gel, le proteine migrano verso il polo positivo a velocità determinate dalle cariche presenti sulle molecole. Trattando poi il gel con coloranti specifici, vengono evidenziate bande caratteristiche. Gli esemplari che mostrano soltanto una banda a rapida o a lenta migrazione sono omozigoti; quelli con due bande sono eterozigoti. Osservando i cambiamenti nella frequenza relativa delle varianti proteiche, si possono misurare, all'interno della popolazione, variazioni di frequenza genica.

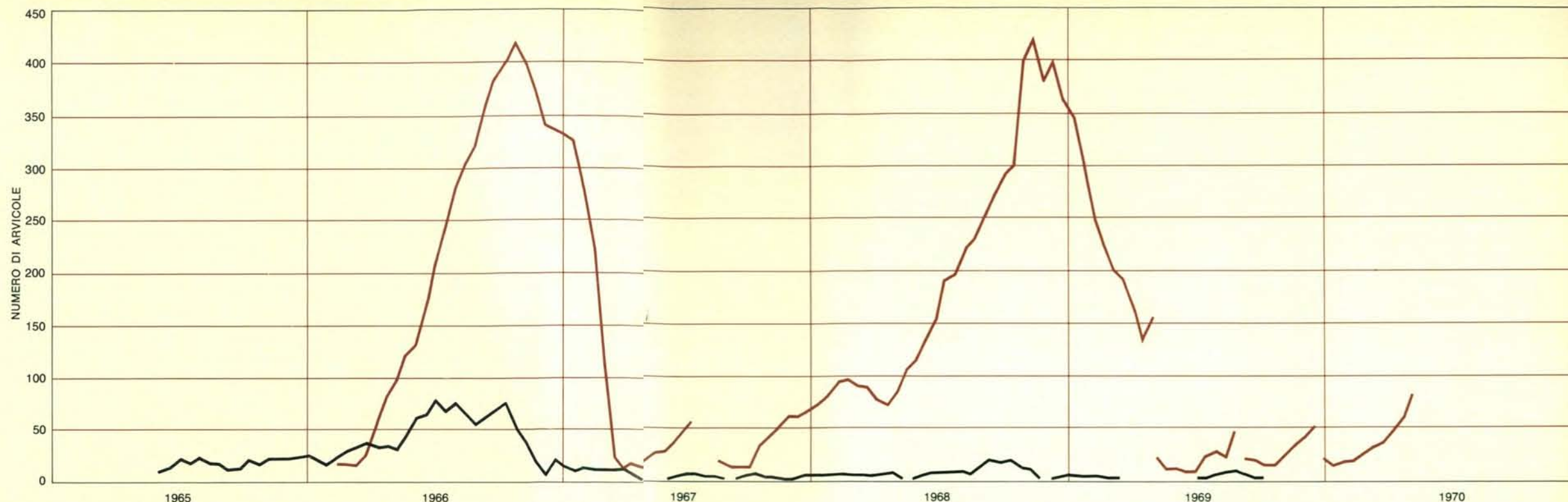
Le variazioni nella frequenza genica dei maschi di *Microtus pennsylvanicus* durante la primavera del 1969 sono state valutate mediante elettroforesi. Le frequenze relative del gene *F* per la leucina-amminopeptidasi (*in nero*) e del gene *E* per la transferrina (*in colore*) sono in stretta correlazione durante tale periodo. I geni non rappresentano caratteristiche note, specifiche degli animali, ma servono semplicemente come comodi indicatori del genotipo. Durante questo periodo la popolazione era in declino.

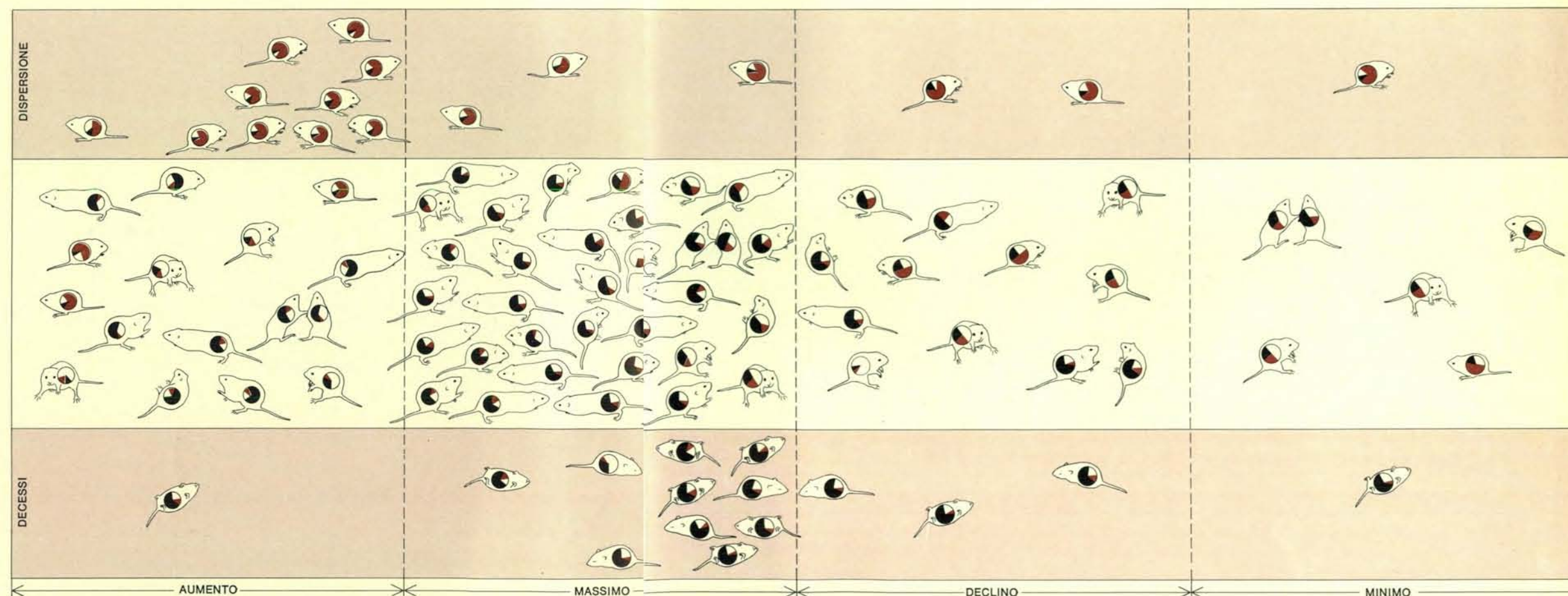
Individui residenti e migratori di *Microtus pennsylvanicus* si distinguono per il genotipo, determinabile attraverso le bande elettroforetiche della transferrina. Tra gli individui che migravano predominavano gli eterozigoti; mentre tra i residenti i genotipi erano più vari, ma l'omozigote Tf^C/Tf^C sembrava il genotipo più comune.

composizione delle proteine non causano i cambiamenti comportamentali osservati nei nostri esperimenti. Si tratta solo di marcatori incidentali con i quali possiamo individuare, nella popolazione, cambiamenti genetici coincidenti con variazioni nei caratteri comportamentali (*si veda l'illustrazione a pagina 58*).

Abbiamo operato principalmente con due sistemi proteici: la transferrina (Tf), una proteina che trasporta il ferro, e l'enzima leucina-amminopeptidasi (LAP), ambedue con due importanti varianti distinguibili mediante elettroforesi nelle popolazioni di roditori che abbiamo studiato. In ambedue le proteine si potevano notare variazioni cicliche che dimostravano corrispondenti variazioni genetiche. Modificazioni particolarmente ampie nella frequenza genica furono osservate nella fase di declino del ciclo della popolazione e la direzione di questo cambiamento fu uguale in parecchi cicli diversi tra loro.

Un cambiamento nella frequenza relativa di un particolare genotipo all'interno di una popolazione può verificarsi attraverso una modificazione del tasso di natalità e di mortalità di individui appartenenti a quel dato genotipo. Nei cicli di popolazione di roditori abbiamo trovato che ambedue i meccanismi sono importanti (*si veda la figura in alto a pagina 59*). In *Microtus pennsylvanicus*, per esempio, lo studio della transferrina ha messo in luce che sia il tasso di natalità, sia il tasso di mortalità di differenti genotipi si modificano sistematicamente lungo il ciclo.

In *Microtus pennsylvanicus* si trovano tre genotipi per la transferrina: l'omozigote con geni che specificano solo per la forma che migra rapidamente (designato come Tf^C/Tf^C), l'omozigote che ha soltanto la forma che migra lentamente (Tf^E/Tf^E), e l'eterozigote che possiede un gene di ogni tipo e mostra nell'elettroforesi ambedue le bande (Tf^C/Tf^E). L'omozigote Tf^C/Tf^C aveva sempre maggiori probabilità di sopravvivere durante l'aumento della popolazione; l'eterozigote Tf^C/Tf^E sopravviveva sempre meglio nei periodi di massimo addensamento, mentre l'altro omozigote Tf^E/Tf^E sembrava essere favorito durante il declino. Queste variazioni nel grado di sopravvivenza furono associate con disparità nella capacità di riproduzione di porzioni della popolazione. Lungo il ciclo i maschi eterozigoti si trovavano più spesso in fase di riproduzione rispetto ai maschi di ambedue i genotipi omozigoti e le femmine che erano o

Popolazioni recintate di *Microtus ochrogaster* non hanno mostrato il caratteristico ciclo di tre-quattro anni; esse infatti si sono diffuse rapidamente, raggiungendo densità cinque volte più elevate di quelle delle vicine popolazioni di controllo. Nuove popolazioni sono state introdotte nell'area per ben cinque volte (*curve in colore*); il fatto che ogni volta sia stata raggiunta la densità massima, conferma l'ipotesi che la dispersione sia una parte importante del meccanismo del ciclo. Anche se la densità massima superava di gran lunga quella della popolazione di controllo (*curva in nero*), le comunità recintate hanno prosperato.

DISPERSIONE
DECESSI
AUMENTO — MASSIMO — DECLINO — MINIMO

Il modello del meccanismo proposto per spiegare il ciclo di popolazione nei roditori è basato su cambiamenti periodici. Arvicole con una tendenza ereditaria verso la dispersione, associata con un'elevata velocità di riproduzione (*aree in colore*), migrano durante il periodo in cui aumenta la popolazione. Nella restante popolazione predominano i geni che determinano l'aggressività (*aree in nero*). L'aggressività è probabilmente un vantaggio adattativo in comunità sovraffollate, ma i roditori aggressivi sono poco adattati nei riguardi di altre caratteristiche (*aree in bianco*); quando il loro tasso di mortalità sale, la popolazione declina.

Sono state create aree recintate di circa un ettaro per la sperimentazione sui roditori vicino all'Indiana University, circondate da una rete metallica e poste in mezzo a fasce di terreno privo di vegetazione. La rete è infissa a 60 cm di profondità per impedire agli animali di attraversarla sottoterra ed è sovrastata da una flangia metallica che gli animali non riescono a superare.

eterozigoti o omozigoti Tf^C/Tf^C erano capaci di una maggior riproduzione rispetto alle omozigoti Tf^E/Tf^E. Infine i genotipi erano distinguibili in base all'età in cui gli animali erano in grado di riprodursi per la prima volta. Quest'età fu calcolata indirettamente misurando il ritmo di crescita dei giovani roditori. Omozigoti Tf^C/Tf^C crescevano più rapidamente di altri durante le fasi di aumento e di massima densità del ciclo e diventavano così sessualmente maturi più presto, mentre nelle popolazioni in declino gli omozigoti Tf^E/Tf^E crescevano più in fretta.

I nostri studi sulla variazione e la selezione naturale delle proteine ci hanno condotto alla seguente interpretazione del ciclo di popolazione di *Microtus pennsylvanicus*. Nel periodo di aumento, l'entità della sopravvivenza è notevole e il vantaggio selettivo dipende da quei genotipi che hanno un elevato tasso di riproduzione, un elevato tasso di crescita e una precoce maturità sessuale. Durante la fase del declino, la mortalità è molto più elevata, particolarmente tra i giovani. Il tasso di riproduzione e quello di crescita sono così meno importanti della pura e semplice sopravvivenza e vengono selezionati quegli individui che hanno avuto successo per la loro resistenza, anche se danno vita a una prole relativamente scarsa.

Questo modello spiega le correlazioni osservate tra variabilità genetica e fluttuazioni nelle popolazioni, ma non indica quale sia la causa e quale l'effetto. L'ipotesi di Chitty presuppone che il ciclo sia conseguenza della trasformazione subita dalla composizione genetica della popolazione, ma questi esperimenti potrebbero dimostrare che i cambiamenti genetici sono un effetto collaterale di un ciclo di popolazione controllato da qualche agente esterno. Per dimostrare l'importanza del meccanismo genetico furono necessari ulteriori studi sul ciclo di popolazione delle arvicole.

In queste ricerche fummo favoriti da un procedimento sperimentale che permette di avviare o arrestare un ciclo a piacere: se una popolazione di roditori è confinata in un recinto, il ciclo si arresta. Questo fatto interessante fu scoperto da John Clarke di Oxford in *Microtus agrestis* e fu confermato in maniera indipendente da A. van Wijngaarden, che stava lavorando allora per il Plant Protection Service di Wageningen in Olanda. Sia Clarke, sia van Wijngaarden usarono recinti relativamente piccoli, aventi all'incirca le dimensioni di una grande stanza. Noi decidemmo invece di studiare gli effetti di recinzioni molto più vaste e, nel 1965, stabilimmo tre aree di prova, ognuna comprendente una superficie di poco meno di un ettaro in un territorio dell'Indiana meridionale. Tali aree vennero circondate da una rete metallica che fu infissa fino a circa 60 cm sotto il tappeto erboso in modo che i roditori non potessero scavare cunicoli sotto di essa; d'altra parte la stessa rete possedeva una flangia di metallo alla sommità perché gli animali non si arrampicassero e passassero oltre *(si veda la figura in questa pagina)*.

L'effetto della recinzione su una popolazione di arvicole fu notevolissimo. All'interno dell'area recintata le popolazioni raggiunsero densità da 4 a 20 volte più elevate di quelle delle vicine popolazioni di controllo. Benché in essa gli animali durante l'inverno provocassero considerevoli danni al loro habitat, a ogni primavera la vegetazione si riprendeva rapidamente. Le velocità di crescita e di sopravvivenza dei singoli individui variano di solito in risposta alle condizioni dell'habitat, ma, a ogni primavera, la crescita rinnovata della vegetazione era accompagnata da un aumento immediato del numero di arvicole. Vennero osservate soltanto fluttuazioni stagionali della popolazione: risultava soppresso il ciclo caratteristico di tre-quattro anni *(si veda l'illustrazione in alto nelle due pagine precedenti)*.

Questi semplici esperimenti rivelarono sulle arvicole meccanismi di regolazione della popolazione che non avrebbero potuto essere scoperti osser-

vando popolazioni non recintate. I risultati suggeriscono immediatamente che lo stress associato a elevate densità non poteva di per sé essere sufficiente a provocare il declino della popolazione. All'interno dei recinti le densità di popolazione raggiunsero livelli molto più elevati di quelli osservati in condizioni naturali e anche lo stress a esse associato fu probabilmente più intenso del normale. Eppure ciò non arrestò la crescita della popolazione. L'esperimento confermò anche che, a densità normali, il deterioramento dell'habitat non poteva essere un fattore limitante per la popolazione.

La crescita sfrenata suggerì un ovvio meccanismo di regolazione della popolazione nelle arvicole: la dispersione.

Era questo processo, più di tutti gli altri, che veniva bloccato dalla recinzione.

Per studiare le caratteristiche degli individui che compivano migrazioni abbiamo analizzato gli animali che si spostavano verso un'area resa da poco disponibile. Per due anni abbiamo eliminato periodicamente dalle due aree a prateria aperta tutte le arvicole e abbiamo reso possibile per due settimane l'immigrazione di altri individui nell'habitat vacante; infine abbiamo tolto anche tutti i nuovi arrivati. Questi, in seguito, furono confrontati con animali che vivevano nelle popolazioni circostanti prese come controlli.

Parecchie caratteristiche distinguevano le arvicole migranti da quelle residenti. Tra le prime vi era un numero sproporzionato di giovani femmine che erano diventate sessualmente mature di recente e mostravano un'evidente notevole tendenza a disperdersi. Di conseguenza, i colonizzatori di un nuovo territorio avevano con tutta probabilità un elevato potenziale riproduttivo. Nel periodo di aumento della popolazione, i genotipi delle femmine migranti di *Microtus pennsylvanicus* differivano da quelli delle femmine residenti, facendo così pensare che lo spostamento di animali da aree sovraffollate ad aree con bassa densità di popolazione avrebbe potuto alterare in maniera significativa la costituzione genetica di ambedue le popolazioni (*si veda la figura in basso a pagina 59*).

La massima incidenza della dispersione fu osservata nei periodi di crescita della popolazione. D'altra parte, nella fase di declino, solo il 15 per cento circa della perdita era dovuta all'emigrazione. Pertanto sembra che la dispersione allontani selettivamente individui che non tollerano il sovraffollamento e abbia un effetto significativo sulla popolazione solo quando la sua densità cresce.

Oggi possiamo costruire un modello del ciclo di popolazione dei roditori che riunisce l'ipotesi di Chitty con i nostri risultati sulla dispersione (*si veda l'illustrazione in basso alle pagine 60 e 61*). Il suo postulato essenziale è che le popolazioni di roditori comprendono due ampi genotipi, uno superiore dal punto di vista della riproduzione, ma intollerante dei forti addensamenti, l'altro adatto alla sopravvivenza in condizioni di sovraffollamento, ma dotato di una velocità di riproduzione più bassa. A mano a mano che la popolazione aumenta, gli individui intolleranti dei forti addensamenti se ne vanno. La probabilità di successo di questi animali non è stata ancora misurata, ma quelli che trovano un habitat adatto non sovraffollato probabilmente se la cavano bene. Invece gli animali che tollerano gli addensamenti elevati rimangono nel loro habitat. Per questa porzione della popolazione l'aggressività, più che il potenziale riproduttivo, è la caratteristica del massimo vantaggio selettivo.

Il declino nella popolazione residente può essere accelerato da un qualsiasi fattore e la variabilità di questa fase del ciclo suggerisce che l'evento istigatore sia diverso nelle varie situazioni e possa essere ampiamente determinato dal caso. Tuttavia, la causa ultima del declino è da ricercarsi quasi certamente nella composizione della popolazione stessa: gli animali lasciati indietro, più aggressivi, ma meno prolifici, sono evidentemente più suscettibili di altri alle normali cause di morte.

Molte di queste conclusioni si basavano su osservazioni effettuate su specie che vivono su una ristretta area geografica; esse perciò non possono venire generalizzate prima che siano stati eseguiti ulteriori esperimenti. Popolazioni formate interamente da animali migrati possono svolgere un ruolo importante nel generare fluttuazioni; eppure è difficile perfino provare che queste esistono. Le caratteristiche degli animali che tendono a compiere migrazioni devono essere studiate in altre specie se si vogliono dimostrare le ipotesi precedentemente enunciate e chiarire i due grossi misteri ancora insoluti: perché il ciclo ha un periodo dai tre ai quattro anni e perché i cicli di popolazioni isolate sono sincronizzati.

Reazioni chimiche rotanti

Alcune reazioni colorate si organizzano visibilmente nello spazio e nel tempo come onde a spirale. Ciò fa pensare che molte altre reazioni chimiche abbiano un aspetto geometrico che deve essere ricercato

di Arthur T. Winfree

Molti di noi sono cresciuti con la convinzione che le reazioni chimiche sono cose che avvengono per combinazione di sostanze in una provetta, che pertanto si tramutano in altre sostanze in modo abbastanza immediato e uniforme. Noi non abbiamo mai visto provette che risplendessero di colori diversi come insegne al neon o che si differenziassero in regioni interdipendenti in conseguenza di reazioni diverse. Naturalmente nessuno dubita che ciò possa accadere in sistemi chimici. Dopo tutto, il mondo è fatto di composti chimici e le reazioni a struttura complessa sono molto comuni. Sembra tuttavia che tali reazioni complicate e interessanti comportino generalmente il trasferimento di calore attraverso la superficie, o un movimento fisico, o membrane che separano materiali diversi, o correnti elettriche che si propagano da una regione all'altra. Ogni scolaro sa che le reazioni dei composti chimici comuni in una soluzione omogenea sono una cosa abbastanza monotona, favorita, nel migliore dei casi, da un singolo cambiamento uniforme nel colore.

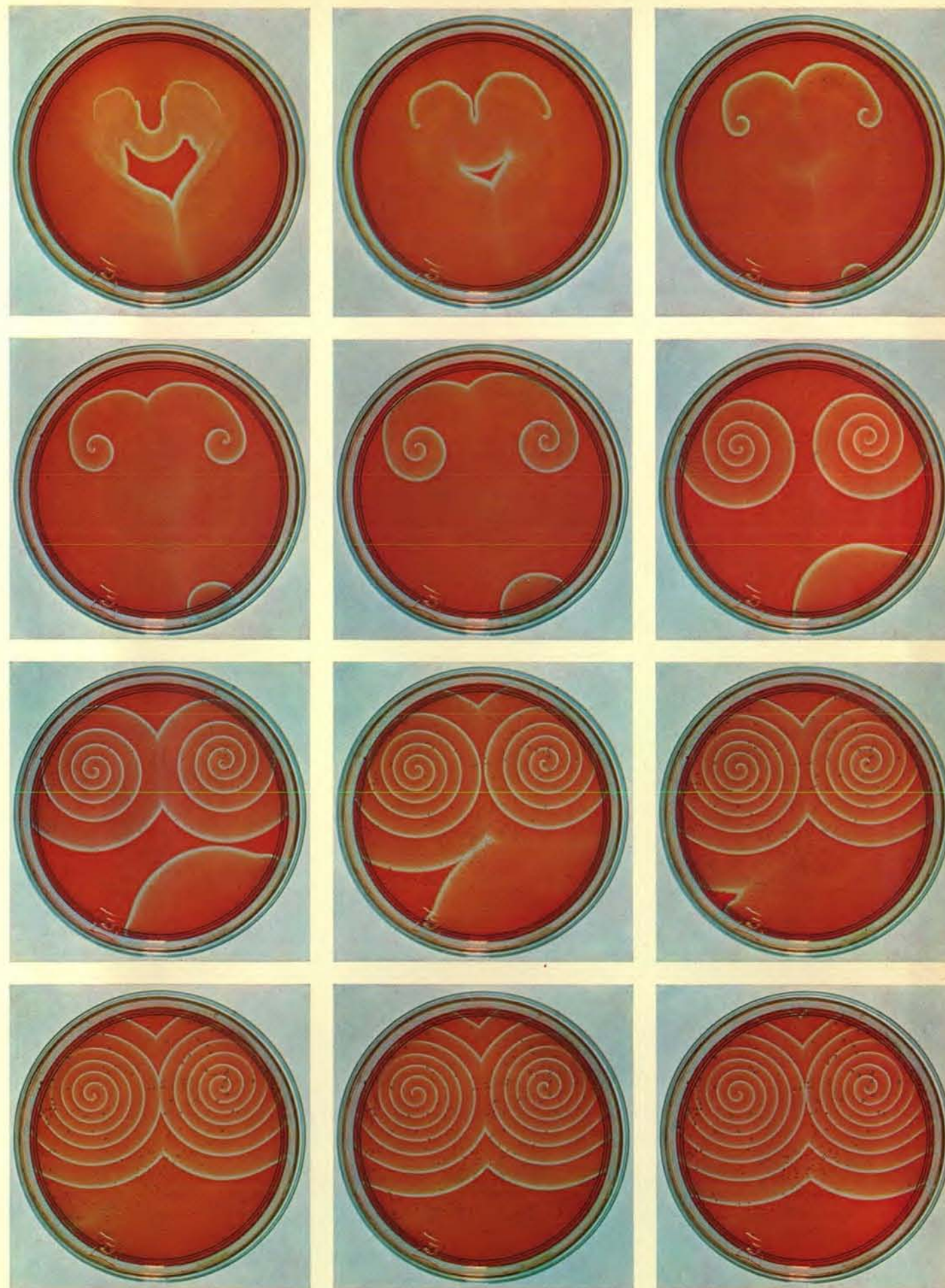
Probabilmente a causa di queste impressioni, A.M. Zhabotinsky e i suoi collaboratori all'Istituto di fisica biologica presso Mosca, si sono impegnati nello studio di una reazione scoperta da B.P. Belousov nel 1958. Nel reagente di Belousov, gli ioni del cerio metallico catalizzano l'ossidazione di un combustibile organico per azione di un bromato in acqua. Diversamente dalla maggioranza delle reazioni comuni, questa, del tutto particolare, non ha uno stato stazionario stabile: essa oscilla con la precisione di un orologio diventando prima gialla, poi incolore, poi di nuovo gialla e così via, cambiando colore due volte al minuto.

Gli esperimenti condotti da Zhabotinsky, benché fossero pubblicati solo in Russia nel corso del 1960, destarono notevole interesse tra i biofisici e i biochimici occidentali che studiavano i fenomeni ritmici negli organismi viventi. Incontrai Zhabotinsky a un congresso a Praga, poche settimane prima dell'occupazione russa della Cecoslovacchia nell'estate del 1968. Io mi sono trastullato con quella reazione abbastanza a lungo per migliorare le mie idee personali sul come si deve studiare l'oscillazione ritmica, nell'arco delle 24 ore, nel cervello del moscerino della frutta, ma nell'eccitamento delle scoperte biologiche, presto dimenticai i reagenti chimici sullo scaffale dei mezzi di coltura.

In seguito, nel 1970, Zhabotinsky e A.N. Zaikin riferirono in un breve articolo apparso su « Nature » che avevano osservato onde chimiche circolari che si propagavano attraverso una versione modificata del reagente oscillante. Queste onde di colore blu intenso emergono periodicamente, formando anelli concentrici simili a un bersaglio, intorno a punti isolati che fungono da *pacemaker* in una capsula poco profonda contenente un liquido rosso-arancio. Quando onde generate da due generatori di impulsi si scontrano, non si compenetrano come le onde in uno stagno, ma svaniscono. Nelle collisioni successive, le onde provenienti da un generatore di impulsi a frequenza più elevata avanzano avvicinandosi passo a passo al generatore di impulsi a frequenza minore. Alla fine, arrivano al *pacemaker* più lento e lo eliminano, riducendo di una unità il numero dei disegni del bersaglio nella capsula.

Qual è la fonte generatrice di impulsi del bersaglio? C'è qualcosa di insolito in quei punti, oppure gli anelli concentrici si formano semplicemente per diffusione e per la cinetica delle reazioni chimiche in un reagente puro? I generatori di impulsi hanno periodi diversi: ciò fa pensare che essi abbiano origine da cause diverse. Generalmente appaiono all'interfaccia tra il liquido e il suo contenitore. Spesso una particella di pulviscolo fluttuante o una scalfittura del vetro della capsula Petri può essere vista al microscopio. Se il liquido è a contatto con ossidanti forti o con alcune leghe metalliche, a quel punto il reagente rosso-arancio diventa blu a intervalli regolari, emettendo onde circolari come un *pacemaker*: ciò suggerisce l'idea che impurità chimiche locali presenti nella soluzione possano essere generatrici di impulsi.

I generatori di impulsi possono essere eliminati filtrando con cura il reagente in una capsula priva di scalfitture e rivestita di silicone. Il reagente così preparato e lasciato in riposo, continua a oscillare quasi omogeneamente. Tale reagente filtrato è ancora in grado di condurre le onde? La risposta è positiva: se una piccola goccia di reagente prelevata dalla parte blu di un'onda di un'altra capsula viene aggiunta alla soluzione omogenea,

Spirali di attività chimica in una capsula di « reagente Z » rosso. Toccando la superficie della soluzione con un filamento caldo è stato indotto un anello blu; quindi la capsula è stata scossa delicatamente per rompere l'anello. Le estremità libere dell'onda circolare si arrotolano attorno a un perno vicino a ciascun punto terminale e, avvolgendosi, formano delle spirali che hanno una distanza uniforme tra le onde. Ciascuna spirale completa ruota con un periodo (τ_0) di circa un minuto. L'ordine della sequenza di foto è da sinistra a destra; le immagini sono state riprese ai seguenti tempi: 0, ½, 1, 1½, 2, 3½, 4½, 5½, 6½, 7, 7½ e 8 minuti. Sono visibili bolle di anidride carbonica che è un prodotto della reazione. La profondità del liquido è di 1,4 millimetri.

Disegni a forma di bersaglio sono formati da anelli di attività chimica nel reagente Z. Essi sono emessi a intervalli diversi, ma regolari, da punti generatori di impulsi che sembrano essere particelle di impurità. Anelli provenienti da generatori più veloci gradualmente predominano su quelli dei generatori più lenti. Le foto sono state scattate a intervalli di mezzo minuto.

Anelli allungati decadono in un paio di spirali simmetriche ad evolvente che si trovano nella posizione delle ore 12. Nella posizione delle 5 una sorgente di anello allungato si scinde con perfetta simmetria. Prima che le sorgenti, che si trovano nella posizione delle 9, delle 12 e delle 2, emettano anelli circolari, svaniscono. Le foto sono riprese a intervalli di 1 minuto.

« infetta » il reagente filtrato e l'infezione si propaga come si espande un solo anello alla velocità standard. Pertanto la mancanza di disegni a forma di bersaglio è dovuta a una mancanza di fonti generatrici di impulsi per onde circolari. Sembra che i generatori di impulsi siano particelle di impurità.

Si verifica anche che, se nella stessa formulazione si ha un po' meno di acido e un po' più di bromuro, l'oscillazione della soluzione può essere eliminata senza peraltro influire sulla sua capacità di condurre le onde di attività chimica. Quello che rimane è una soluzione stabile rosso-arancio che io chiamerò reagente Z in onore di Zhabotinsky e Zaikin. Il reagente Z non dà luogo ad alcun fenomeno quando è lasciato solo; ma quando è stimolato per l'aggiunta di una piccola goccia « infettante » di onda blu presa da un'altra capsula, o perché toccato con un filamento caldo, un singolo anello blu intenso si propaga attraverso di esso alla velocità costante di pochi millimetri al minuto.

I dettagli del meccanismo della reazione e della cinetica nel reagente Z stanno per essere chiariti solo ora. Elaborando gli studi iniziali intrapresi da Zhabotinsky, Zaikin e da V.A. Vavilin, Richard J. Field e Richard M. Noyes dell'Università dell'Oregon e E. Körös a Budapest hanno di recente proposto uno schema per il reagente di Belousov che comporta reazioni tra 11 sostanze. Field e Noyes hanno inoltre dimostrato che il loro schema può essere spiegato con una sequenza di reazioni che implica solo tre sostanze, che oscillerebbero spontaneamente.

La loro sequenza di reazioni si comporta pressappoco così. I costituenti di base della soluzione sono bromuri e bromati (due tipi di composti del bromo), malonato (un combustibile organico) e la fenantrolina (un indicatore colorato contenente ferro). Il suo colore è rosso quando l'atomo di ferro è ridotto a ione ferroso con due elettroni nella sua orbita esterna, mentre è blu quando è ossidato a ione ferrico, con tre elettroni nell'orbita esterna. La concentrazione degli ioni bromuro nella soluzione determina quale dei due gruppi di reazioni sarà dominante. Il primo gruppo è associato a una concentrazione elevata di ioni bromuro e alla presenza dello ione ferroso; la soluzione, pertanto, è rossa. Il secondo gruppo di reazioni è inibito dal bromuro ed è associato alla forma ossidata (ione ferrico); la soluzione è blu.

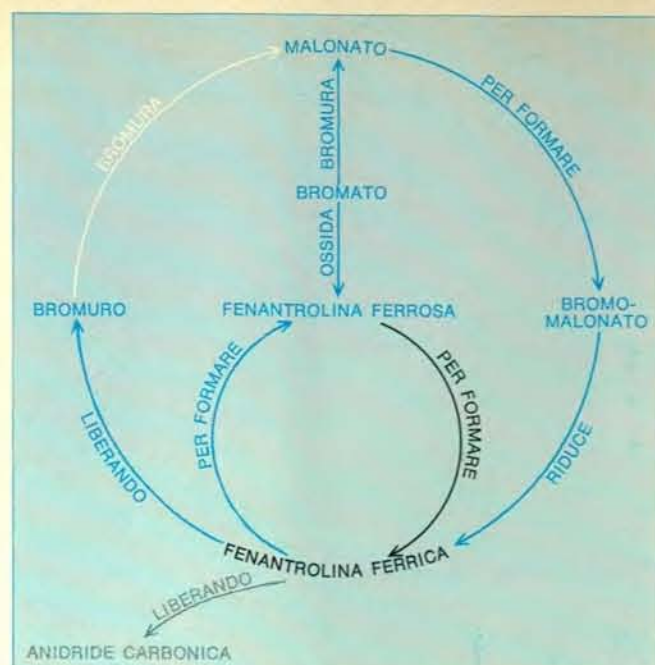
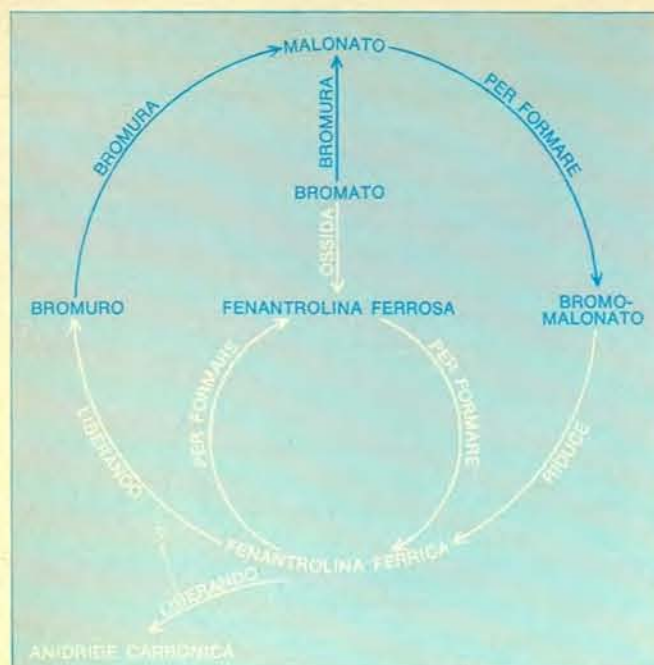
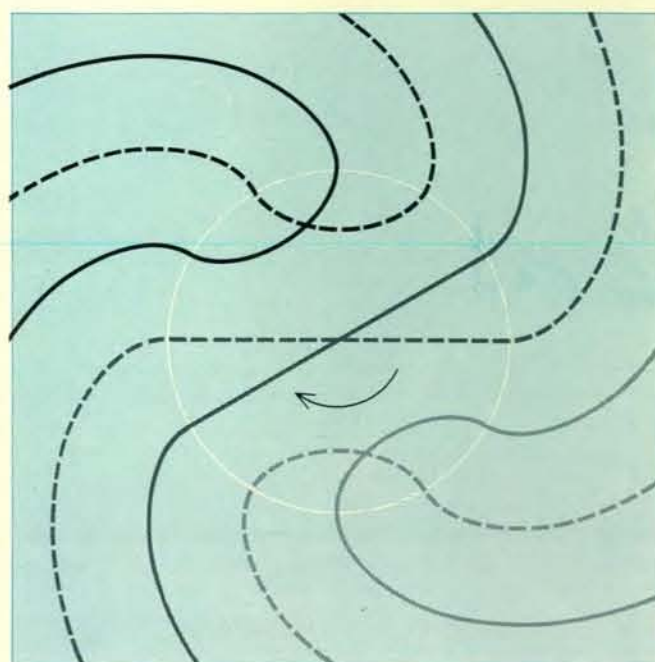
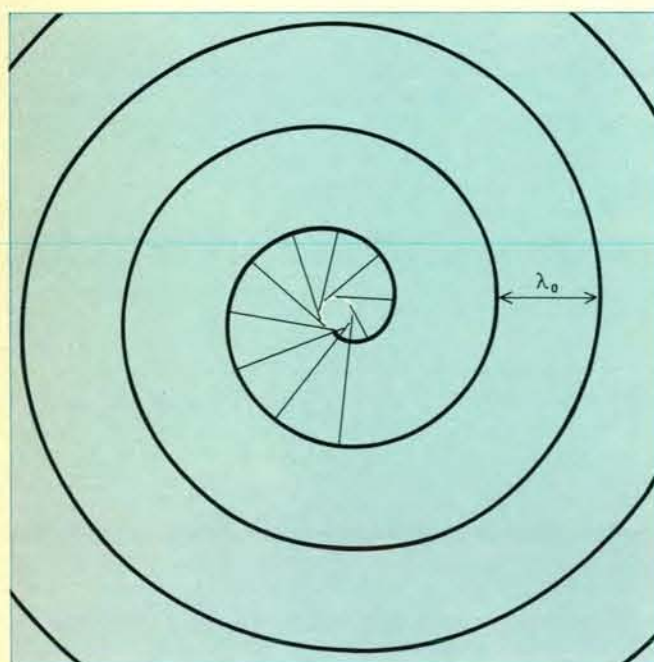
Quando la concentrazione del bromuro è elevata, questo viene consumato nel primo gruppo di reazioni per bromurare il malonato con formazione del bromomalonato. Quando la concentrazione del bromuro scende al livello critico, incomincia a predominare il secondo gruppo di reazioni e vengono consumate le ultime tracce di bromuro. Nel frattempo, il bromato opera la bromurazione del malonato; contemporaneamente ossida il ferro contenuto nella molecola dell'indicatore, trasformandolo da ione ferroso a ferrico, così che il colore del complesso della fenantrolina cambia da rosso a blu. La fenantrolina ferrica ossida quindi il bromomalonato che si è formato. Questa reazione libera bolle di anidride carbonica e bromuro in misura tale che la seconda sequenza di reazione cessa e riprende nuovamente la prima (*si veda l'illustrazione in alto a pagina 74*).

Reazioni analoghe governano probabilmente l'eccitabilità del reagente Z che non oscilla, preparato con un eccesso di ioni bromuro e con una minor quantità di acido. In questo caso le reazioni rosse e blu si compenseranno con la concentrazione degli ioni bromuro proprio al di sopra del livello critico: il rosso domina finché interviene qualche fattore esterno. Seguendo tale perturbazione, il fronte d'onda blu si propaga consumando le piccole quantità di bromuro che si diffondono dal liquido rosso adiacente, permettendogli così quasi di volgere nella fase blu. Una volta deviate, le reazioni ritornano alla fase rossa in attesa che scatti il prossimo fattore di perturbazione.

Poiché molti studiosi di tutto il mondo hanno mostrato un notevole interesse a queste reazioni, sembra probabile che i processi chimici saranno presto spiegati con dati quantitativi completi, rendendo così possibile l'ottenimento di varianti più interessanti. Non è necessario nulla di più sofisticato, tuttavia, se tutto ciò che si desidera è di avere a che fare con strani fenomeni, mentre si pensa alle reazioni spaziali che si organizzano da sole. Il fenomeno che mi ha affascinato è la creazione di onde a spirale di attività chimica nel reagente Z.

Segmenti di un fronte circolare di onde che si propagano attraverso il reagente possono essere fatti svanire scuotendo leggermente la capsula. Le restanti onde frammentarie incominciano ad arricciarsi attorno a un perno vicino a ciascuna estremità, avvolgendosi alla fine in spirali, che hanno una distanza tra le onde di circa due millimetri. L'intero disegno ruota una volta ogni 20-60 secondi, a seconda della formulazione (*si veda l'illustrazione a pagina 71*).

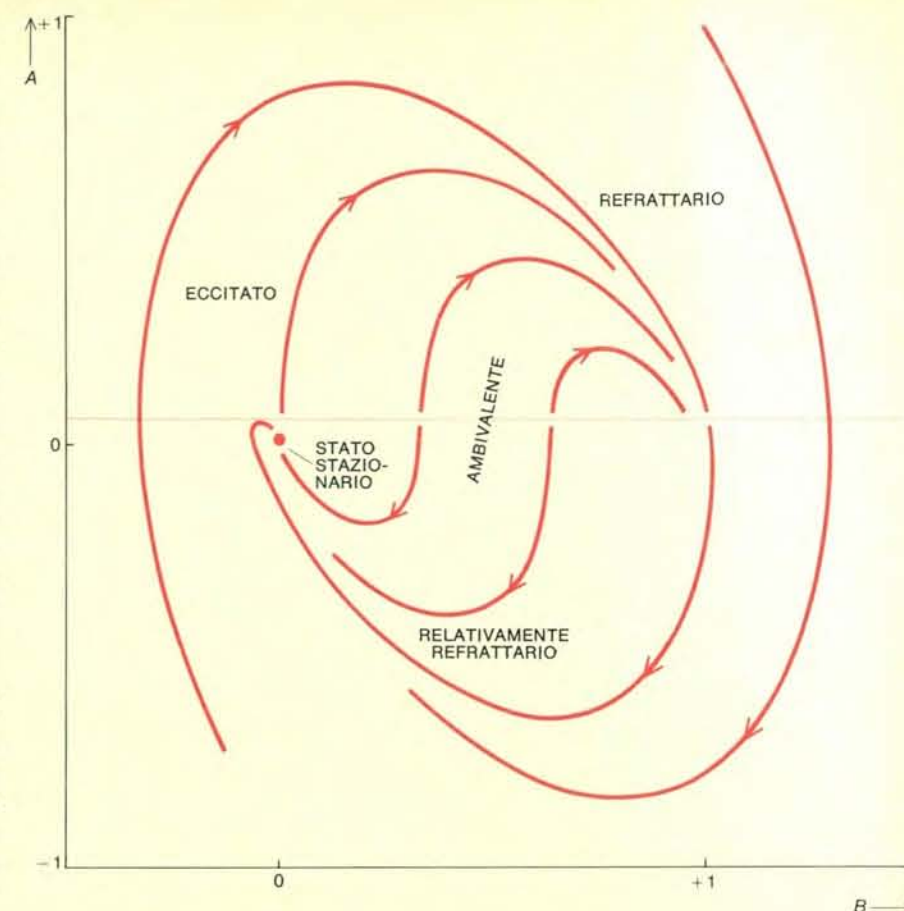
Una nota di Zhabotinsky rivelò che anch'egli aveva incontrato spirali. Infatti, matematici e biofisici russi, nel tentativo di capire le vibrazioni e le fibrillazioni nel cuore, avevano studiato le onde a spirale in mezzi eccitabili alcuni anni prima che il mezzo chimico venisse scoperto. I ricercatori russi incominciarono a interessarsi al problema leggendo la pubblicazione, apparsa nel 1946, del matematico Norbert Wiener e del cardiologo Arturo Rosenblueth: *Archives of the Mexican Institute of Cardiology*. Essi esaminarono un modello di onde elettricamente eccitate che giravano attorno a ostacoli nel muscolo cardiaco. Le loro prime analisi furono elaborate con metodi sia matematici sia biologici negli anni '60, in buona parte attraverso le osservazioni di I.S. Balakovskii e V.I. Krinskii sulle onde a spirale. Queste analisi potrebbero fornire chiarimenti sulla organizzazione dinamica delle onde a spirale di una reazione chimica in soluzione?

In un certo senso, sì. Questa linea di analisi, tuttavia, si basa sul concetto che lo stato di una cellula eccitabile cambia solamente in un modo importante: può trovarsi in qualsiasi luogo in un continuo di stati varianti dallo stato di quiete al « refrattario » attraverso lo stato eccitato e di nuovo allo stato di quiete. Questa può essere una descrizione chiara delle cellule biologiche e forse anche della propagazione delle onde nel reagente Z. Quando ci si riduce a una sola variabile, tuttavia, questa interpretazione impedisce la ricerca nell'azione reciproca della diffusione e delle reazioni che realmente generano onde nel centro rotante di una spirale. Ogni modello che si basa solamente su una singola variabile richiede che ci debba essere un gradiente di concentrazione infinitamente elevato nel centro di un'onda rotante. Per-

Due tipi di reazioni possono render conto dell'oscillazione del reagente Z da rosso a blu e di nuovo a rosso. Il meccanismo può essere spiegato in forma schematica descrivendo le reazioni tra il bromuro, il bromato, il malonato e la fenantrolina che funge da catalizzatore e da indicatore. La concentrazione del bromuro determina quale dei due tipi di reazione dominerà in una certa regione. Nel primo tipo (*a sinistra*), il bromuro e il bromato bromurano il malonato per formare bromomalonato. Durante questo processo la fenantrolina è rossa, perché il suo atomo di ferro è ridotto a ione ferroso. Se la concentrazione del bromuro cade sotto il livello di soglia, allora ha il sopravvento il secondo tipo di reazione (*a destra*). Quando il bromuro è consumato, è il bromato che opera la bromurazione del malonato. Contemporaneamente ossida l'atomo di ferro trasformando la fenantrolina ferrosa rossa in fenantrolina ferrica blu. Il bromomalonato che si è accumulato riduce quindi la fenantrolina ferrica alla forma ferrosa rossa, liberando bromuro e anidride carbonica. Una elevata concentrazione di bromuro interrompe questa sequenza di reazioni. In questo ciclo sono coinvolte altre sostanze, come l'ossigeno, non indicate nello schema.

Le spirali ad evolvente nel reagente Z hanno proprietà geometriche caratteristiche. Ogni linea tracciata perpendicolarmente all'onda a spirale che avanza (*a sinistra*) e prolungata verso il centro della spirale sarà tangente a un piccolo cerchio (*bianco*), che delimita il nucleo della spirale e la cui circonferenza ha la stessa lunghezza della distanza λ_0 tra le onde della spirale. Il fronte d'onda visibile è una « isobara di concentrazione », cioè una linea di contorno di concentrazione chimica uniforme entro la soluzione. Al di fuori del nucleo della spirale tutte le isobare di concentrazione sono spirali ad evolvente e le isobare di sostanze diverse corrono parallelamente l'una all'altra. All'interno del nucleo (*a destra*), invece, le isobare si intersecano tra loro. Sono illustrate isobare di due reagenti diversi: *A* (*linee intere*) e *B* (*linee tratteggiate*). Le linee nere rappresentano una concentrazione molto elevata, quelle grige scure una concentrazione minore e quelle grige chiare la concentrazione più bassa.

tanto noi dobbiamo avere un modello che comprende due variabili.

Per ragioni di semplicità non iniziamo con una spirale allo stato iniziale, ma con una spirale già formata e simmetrica. Se si accentra l'attenzione su una piccola sezione della spirale, si può così descrivere l'onda vicina a quella sezione: (*a*) « Ogni segmento del fronte dell'onda si propaga perpendicolarmente alla sua lunghezza quasi alla stessa velocità, v_0 ». Una descrizione globale della spirale può essere la seguente: (*b*) « L'onda ruota rigidamente con un periodo τ_0, come se fosse incisa su un disco fonografico che gira ». Combinando la descrizione locale (*a*) con quella globale (*b*) secondo considerazioni geometriche, si può derivare che la distanza perpendicolare, λ_0, tra spire adiacenti della spirale deve essere la stessa in ogni punto, vale a dire $v_0 \tau_0$. Inoltre, il prolungamento interno di ogni linea tracciata perpendicolarmente all'onda a spirale sarebbe tangente a un piccolo cerchio, di circonferenza λ_0, vicino al centro (*si veda l'illustrazione in basso nella pagina a fianco*). L'onda a spirale è l'evolvente di questo cerchio: la curva tracciata da una penna legata al cerchio con una corda tesa e che ruota attorno a esso. Se la distanza tra le onde, λ_0, è poco più di due millimetri, il cerchio ha un diametro di un terzo o un quarto di millimetro.

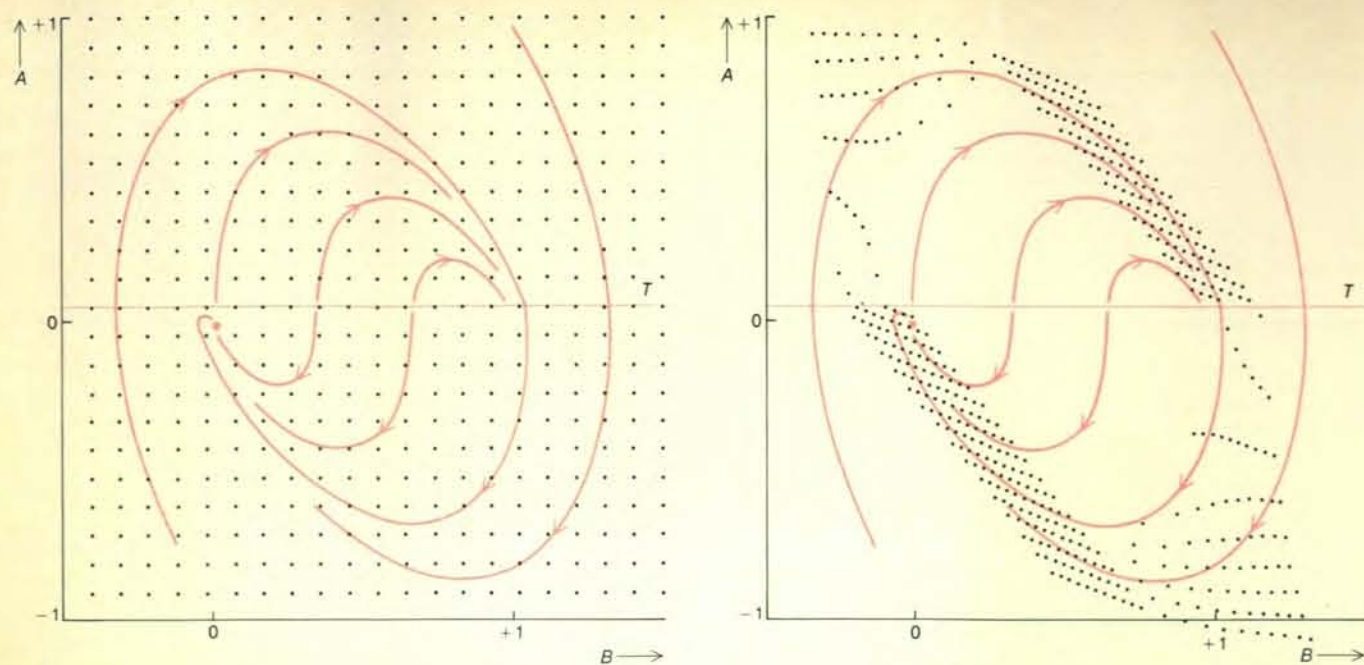
Proprio davanti al fronte dell'onda, l'indicatore fenantrolina è rosso in modo uniforme (forma ridotta); dietro al fronte dell'onda è blu (forma ossidata) e gradualmente cambia fino a diventare di nuovo rosso. Al fronte dell'onda stesso, qualche sostanza (probabilmente lo ione bromuro) passa apparentemente attraverso una concentrazione critica che sposta l'equilibrio dell'ossido-riduzione. Il fronte visibile dell'onda è il luogo geometrico dei punti di una concentrazione particolare; lo chiamerò « isobara » di concentrazione, adottando il termine dei meteorologi, che indica un luogo di pressione uniforme su una carta meteorologica. Ogni elemento di volume, o ogni piccolo volume compatto di reagente, lungo l'isobara ha la stessa concentrazione. Anziché il termine elemento di volume, userò, per brevità, « glob » che significa un volume così piccolo che può essere considerata nulla qualsiasi differenza nella concentrazione dei reagenti.

Se è vero che questi piccoli glob a ogni distanza dietro il fronte visibile dell'onda hanno quasi la stessa composizione chimica, ne deriva che le isobare di concentrazione di tutte le sostanze in gioco devono anche essere evolventi del piccolo cerchio centrale.

Lo spazio di concentrazione è un espediente utile per illustrare la cinetica di reazione di *A* e di *B*. Le coordinate di posizione (*A*, *B*) di ciascun punto sul diagramma rappresentano una possibile composizione (*A*, *B*) di un « glob » nel mezzo eccitato. Sulle ordinate, la concentrazione di *A*, sulle ascisse la concentrazione di *B*. Le equazioni di reazione specificano come *A* e *B* cambino da una data serie di valori iniziali e pertanto specificano una traiettoria (*frecce in colore*) attraverso ciascun punto (*A*, *B*). Nell'illustrazione *B* cambia ad una velocità *KA* (dove *K* è una costante di formazione) e *A* cambia alla velocità di $-(A+B)$. Lo stato stazionario si trova al punto in cui $A = B = 0$ (*punto in colore*). Oltre la soglia *T* in una regione eccitata, la velocità di cambiamento di *A* diventa $-(A+B)+1$. La concentrazione di *A* aumenta rapidamente per un momento, prima che tutte le traiettorie ritornino allo stato stazionario.

Poiché le isobare di concentrazione ruotano attorno al cerchio, ogni glob lungo un'isobara particolare passa attraverso lo stesso ciclo di variazione nella composizione: rosso, blu, rosso, blu, e così via.

Ma qui è evidente un paradosso: è impossibile che glob vicini al centro di un modello di concentrazione rotante passino attraverso lo stesso ciclo di variazioni in ciascuna rotazione. Se lo facessero, significherebbe che ci sono gradienti di concentrazione elevati vicino al perno, situazione che non potrebbe persistere a lungo a causa della diffusione molecolare.

Questo paradosso discende da quello a cui mi riferivo prima dicendo che un modello chimico comprendente una sola variabile non basta a spiegare un'onda a spirale. Entro il cerchio magico centrale, nel nucleo della spirale, la descrizione locale della spirale (*a*) e quella globale (*b*) sono incompatibili. In termini matematici, l'equazione differenziale generata da queste due descrizioni non ha soluzioni che distano meno di $\lambda_0/2\pi$ dal centro della spirale. E che cosa accade delle isobare di concentrazione in quella regione? Esse non possono proprio terminare. Ogni glob ha una certa composizione chimica, e la composizione della soluzione deve variare in modo continuo nello spazio.

Supponiamo che sia possibile compiere un giro d'ispezione in senso orario lungo il contorno del nucleo della spirale. Procederemmo in ordine attraverso glob in ogni fase del ciclo d'onda, dallo scattare del rosso al blu, al ritornare del blu al rosso. Poiché per ogni sostanza in gioco « ciò che sale deve scendere », dobbiamo incontrare ogni isobara di concentrazione una volta quando procediamo dalla bassa verso l'alta concentrazione nel nostro giro in senso orario, e una volta quando

La sola reazione chimica spinge i « glob » (*puntini neri*) lungo traiettorie attraverso lo spazio di concentrazione. Qui sono rappresentati 441 glob con composizioni iniziali diverse al tempo zero (*a sinistra*). (Il reticolo corrisponde a una piccola area del mezzo, di 0,7 millimetri di lato.) Ciascun glob segue indipendentemente la sua traiettoria (*a destra*), poiché nel calcolo fu omessa la diffusione tra glob adiacenti. Sette secondi e mezzo più tardi gli stessi si stanno accumulando nello stato stazionario.

La sola diffusione, senza la reazione, è iniziata per gli stessi 441 glob con le stesse composizioni iniziali, al tempo zero, di quelli della figura della pagina a fianco (*a sinistra*). La differenza consiste nel fatto che i glob sono collegati tra loro a causa della diffusione molecolare, che tende a uguagliare le concentrazioni di A e di B nei glob adiacenti, cosicché essi si muovono insieme nello spazio di concentrazione. Così il reticolo si contrae, (*a destra*) sette secondi e mezzo più tardi.

lo ripercorriamo in senso contrario dall'alta alla bassa concentrazione. Questi due rami della stessa isobara devono in qualche modo connettersi armonicamente entro il nucleo della spirale (*si veda l'illustrazione in basso a pagina 74*).

Queste isobare, tuttavia, si muovono: l'intero disegno ruota rigidamente attorno al nucleo della spirale. Perciò la concentrazione è crescente nei glob su uno dei rami e decrescente nei glob sull'altro. Un tale comportamento lo si può spiegare solo se una qualche seconda sostanza ha una concentrazione diversa nei due rami di una singola isobara. Così la seconda sostanza, sebbene anche le sue isobare siano evolventi che si congiungono entro il nucleo della spirale, deve avere un diverso comportamento nei riguardi dell'aumento e della diminuzione di concentrazione lungo il contorno circolare del nucleo. Il risultato è che, dentro al nucleo, le isobare di concentrazione non sono parallele come fuori dal nucleo: esse si intersecano tra loro.

Questa è la caratteristica critica del nucleo che lo distingue dalle regioni esterne della spirale. Queste ultime propagano puramente la perturbazione creata dalla rotazione del nucleo. Lungo le isobare al di fuori del nucleo, la composizione di ogni glob può essere specificata mediante un unico numero, qualcosa di simile alla fase nel ciclo di stati di Wiener, dallo stato di riposo, all'eccitato, al refrattario, allo stato di riposo. Entro il nucleo la composizione di ogni glob sembra variare indipendentemente in due modi separati. Se ciò è realmente quanto accade, allora causando artificialmente gradienti di concentrazione incrociati nel reagente Z, sarebbe possibile creare un nucleo a spirale, che ruotando emetta un'onda a spirale. E sperimentalmente ciò sembra essere vero.

Ma perché è vero? Quali sono i processi di base che causano questo modo pittoresco di organizzazione spaziale e temporale in una reazione che è anche capace di ossidare semplicemente il malonato in uno stato stazionario lento e uniformemente rosso? Ci sono due fattori che collaborano a generare le spirali, ognuno dei quali di per sé si risolverebbe soltanto in uno stato stazionario omogeneo. Uno è il processo puramente fisico di diffusione molecolare; l'altro è il processo puramente chimico di reazione. I due fattori combinati danno luogo a una attività viva.

Dapprima consideriamo da vicino la diffusione. Che cosa accade con esattezza? A livello molecolare la diffusione è la conseguenza dello spostamento di particelle da una regione affollata a una meno affollata, più spesso che nel senso opposto, semplicemente perché c'è un numero minore di particelle che procede dalla regione meno affollata. La diffusione spinge a scambiare tra loro le concentrazioni delle sostanze in due glob adiacenti della soluzione a una velocità proporzionale alla loro differenza di concentrazione. Il processo si arresta solo quando non rimane alcuna differenza. Nel mondo delle concentrazioni la diffusione è il grande equilibratore per la realizzazione di strutture spaziali nelle soluzioni chimiche.

Consideriamo la reazione, ossia la conversione di popolazioni di molecole in collisione da una forma chimica a un'altra. Dimentichiamoci dello spazio fisico, dei gradienti di concentrazione e della diffusione; pensiamo invece a un glob isolato, omogeneo. Entro tale glob le concentrazioni variano a velocità determinate attraverso le equazioni di cinetica di reazione. Che genere di cinetica di reazione caratterizza un mezzo eccitabile?

Un'analisi di modelli quantitativi tramite il calcolatore mostra che il comportamento delle onde a spirale è quasi indipendente dai dettagli chimici. Invece esso deriva in larga misura dalle caratteristiche qualitative dell'eccitabilità. Adottiamo allora il più semplice modello a due variabili della dinamica di una reazione eccitabile. Siano A e B le concentrazioni di due sostanze, misurate come deviazioni positive o negative dalle loro rispettive concentrazioni allo stato stazionario. Postuliamo la seguente cinetica: B varia alla velocità KA, dove K è una costante di formazione, e A varia alla velocità $-(A + B)$. Così allo stato stazionario A e B sono zero, come è richiesto dalla nostra definizione. Quando A, però, supera una concentrazione critica T, la sua velocità di trasformazione diventa $-(A + B) + 1$. Questa convenzione fornisce l'eccitabilità necessaria.

Se K è abbastanza grande, la corrispondente equazione si comporta come un orologio meccanico, con A e B associati alla posizione e al momento dello scappamento. Finché A è spinto sopra T («tic») nulla accade; poi ritorna giù (« toc ») e oltrepassa il punto di equilibrio oltre T di nuovo (« tic, toc, tic... »). Alternativamente, se K è un po' più piccolo e A è interpretato come potenziale di membrana in una cellula vivente, allora l'equazione è una versione derivata delle equazioni differenziali complesse che descrivono l'eccitabilità del tessuto nervoso e muscolare, equazioni per le quali Alan L. Hodgkin e Andrew F. Huxley meritarono un premio Nobel nel 1963.

Nei nervi, « tic » è seguito solo da « toc » e poi da un ritorno allo stato stazionario. L'equazione di velocità per A può essere arricchita di un termine rappresentante la « diffusione di potenziale » da regioni vicine lungo una fibra lunga di un tale mezzo eccitabile. In questo modo John Rinzel e Joseph Keller del Courant Institute of Mathematical Sciences di New York hanno ottenuto soluzioni analitiche per le caratteristiche essenziali di un impulso a « tic-toc » che si propaga in una fibra nervosa unidimensionale.

Un'implicazione importante per il nostro caso, tuttavia, è che se A e B vengono interpretate come le concentrazioni di due sostanze soggette a diffusione a uguali velocità, la stessa equazione riproduce l'eccitabilità del reagente Z.

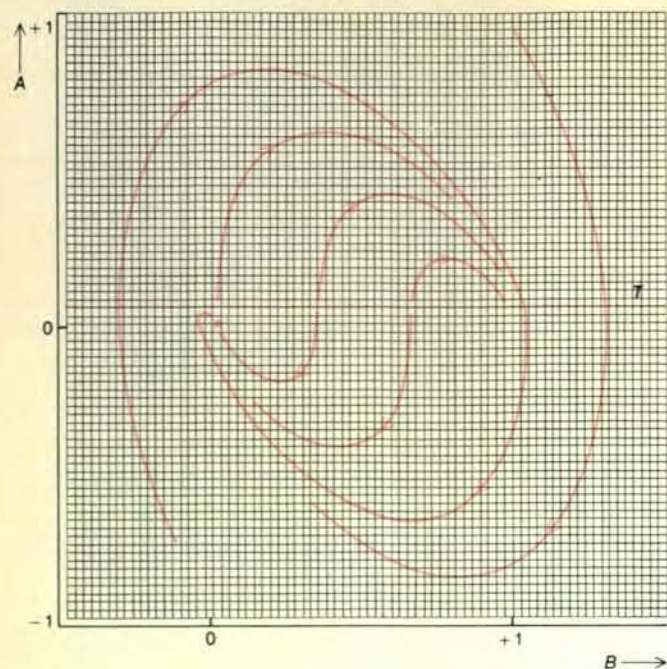
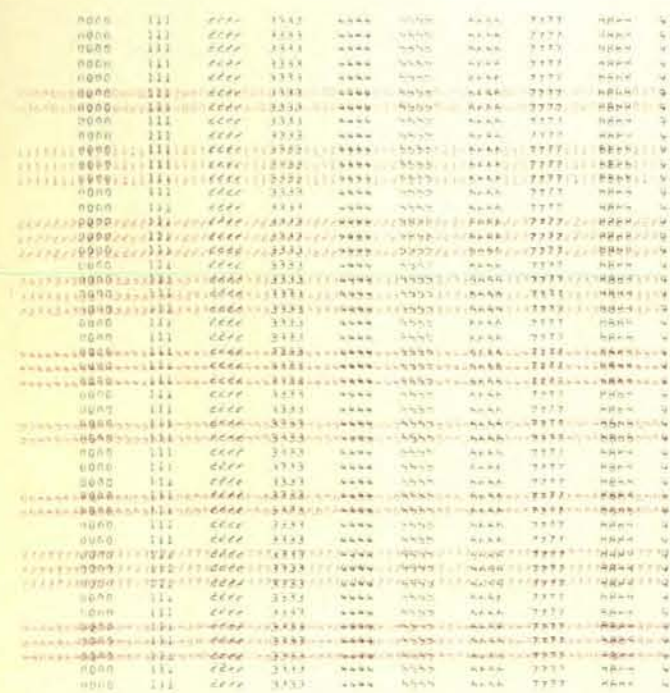
Ora esercitiamo la nostra intuizione geometrica su un modello figurato di queste equazioni cinetiche. La trovata essenziale è di lasciare che le coordinate A e B rappresentino le coordinate di posizione di un minuscolo glob di reagente in un immaginario « spazio di concentrazione ». Ogni punto (A, B) in questo spazio rappresenta una possibile composizione del glob quale è definito dalle sue coordinate di concentrazione A e B (*si veda l'illustrazione a pagina 75*).

La composizione chimica della soluzione varia per il fatto che A e B vengono sintetizzate e distrutte in reazioni le cui velocità sono governate dalle concentrazioni locali sia di A, sia di B. In termini geometrici, si può immaginare una piccola freccia posta in ciascun punto dello spazio di concentrazione che indica come A e B dovrebbero variare insieme nell'istante successivo. Seguendo le frecce si può osservare, con un'occhiata fulminea, l'intera gamma dei possibili modi di comportamento nel glob isolato. Partendo da una qualsiasi composizione iniziale, il futuro delle sue concentrazioni di A e di B si svolge lungo traiettorie nello spazio di concentrazione.

Seguendo le frecce della figura a pagina 75, si può osservare come A e B mutino simultaneamente, partendo da una qualsiasi composizione iniziale, in un glob isolato e omogeneo di un reagente eccitabile. È evidente che « tutte le strade conducono a Roma » ossia allo stato stazionario nel quale entrambe, A e B, sono uguali a zero. Proprio al di là della soglia T, tuttavia, le strade compiono un'ampia deviazione a ritroso. Questa trascina un glob in una fase eccitata ad alta concentrazione di A. Il risultato è contagioso, in quanto la diffusione di molecole di A entro un glob adiacente in quiete porterà anche la sua concentrazione di A al di sopra della soglia.

Mentre il glob è al di sopra della soglia, la sintesi di B, essendo proporzionale alla concentrazione di A, procede più veloce del solito. Man mano B si accumula e la velocità con cui A decade aumenta. La concentrazione di A si abbassa in una fase refrattaria nella quale un successivo contatto con glob eccitati non può far molto per impedire l'inevitabile ritorno allo stato stazionario.

La repentina transizione di colore nel reagente Z dal rosso al blu scatta apparentemente allorché la concentrazione di A va oltre la soglia. Il reagente, la cui composizione è prossima allo stato stazionario, è rosso.

Per poter capire la struttura di un'onda rotante si devono seguire le distribuzioni spaziali di A e B. Immaginiamo ciascun minuscolo glob di reagente in una piccola capsula quadrata rappresentata nello spazio di concen-

La reazione e la diffusione collaborano a rafforzare l'attività in un mezzo eccitabile. Il grafico dell'elaboratore indica ciò che succede in una sezione del mezzo di 35 mm², usando un fine reticolo di 5625 glob nello spazio di concentrazione. Inizialmente i glob sono localizzati in modo da ottenere gradienti di concentrazione incrociati (*a sinistra*). Dopo 315 secondi, il reticolo ruota stabilmente. I glob sul bordo raggiungono lo stato stazionario, ma poi vengono spinti al di là della soglia per effetto della diffusione. I glob nel centro del reticolo sono spinti attraverso la regione ambivalente (*a destra*) e variano di poco.

La simulazione da parte dell'elaboratore presenta, nello spazio fisico reale, i risultati congiunti della reazione chimica e della diffusione nell'immaginario spazio di concentrazione (*si veda l'illustrazione in alto*). Le isobare di concentrazione iniziale di *A* e *B* sono tracciate perpendicolarmente tra loro in una capsula quadrata del mezzo eccitabile. Le concentrazioni sono indicate con numeri da 0 a 9. La concentrazione di *A* (*numeri in nero*) aumenta da sinistra a destra nella capsula, la concentrazione di *B* (*numeri in colore*) dall'alto verso il basso.

Forme d'onda a spirale, nella simulazione dell'elaboratore, si possono qui osservare dopo un tempo di 315 secondi. La concentrazione di *A* è indicata da numeri neri e quella di *B* da numeri in colore. La linea nera indica dove *A* sta aumentando attraverso la soglia e dove il repentino fronte d'onda blu si troverebbe nel reagente Z. Le isobare di concentrazione di *A* e *B* appaiono parallele esternamente al nucleo della spirale (*cerchio in colore*), mentre all'interno si incrociano. I successivi stampati dell'elaboratore sono simili, ad eccezione di una rotazione.

trazione: la composizione (*A,B*) di ciascun glob nella capsula determina le sue coordinate (*A,B*) nello spazio di concentrazione. Poiché i glob adiacenti hanno una composizione simile, la rappresentazione risulta scorrevole, senza vuoti o discontinuità. Supponiamo che la concentrazione di *B* aumenti dall'alto verso il basso nella capsula, e che *A* aumenti da sinistra a destra, in modo che esistano gradienti di concentrazione che si incrociano in composizioni ambivalenti nel centro della capsula. La cinetica della reazione chimica muove adesso ciascun glob (non fisicamente, ma nello spazio di concentrazione) lungo una certa traiettoria. Il movimento è generalmente nel senso orario attorno alla regione ambivalente centrale, ad eccezione di una ristretta gamma di composizioni comprese tra lo stato stazionario (dove entrambe *A* e *B* sono uguali a zero) e la soglia *T*. In detta regione le frecce sono dirette indietro in senso antiorario verso lo stato stabile stazionario.

Se non vi fosse diffusione, i glob al di sopra della soglia continuerebbero a spostarsi attorno nel senso orario, mentre quelli appena al di sotto della soglia ricadrebbero nello stato stazionario per effetto di una decrescente concentrazione di *A*. Invece, i glob inizialmente appena al di sotto della soglia hanno il loro contenuto di *A* incrementato per diffusione da parte dei loro vicini, attualmente ben al di sopra della soglia; così essi vengono spinti loro malgrado sopra la soglia. Una volta verificatosi questo fatto, essi renderanno lo stesso servizio ai loro vicini, e in tal modo il processo continua. Come una macchina a moto perpetuo, la capsula di reagente ruota attorno alla regione ambivalente nello spazio di concentrazione.

La diffusione di molecole tra glob adiacenti conferisce al reagente una specie di coerenza nello spazio di concentrazione che agisce nel senso di livellare le variazioni locali delle velocità di reazione: la piccola gamma di composizioni nella quale glob isolati slittano indietro verso lo stato stazionario viene appianata facendo prevalere per gli stessi la generale tendenza a ruotare nello spazio di concentrazione. Inoltre, nello stesso tempo in cui l'immagine del reagente ruota nello spazio di concentrazione, le isobare di concentrazione ruotano fisicamente nel reagente.

Questa situazione si verifica soltanto se sono inizialmente disponibili gradienti di concentrazione convenientemente elevati. Se il reagente è inizialmente pressoché omogeneo entro tutta la capsula, allora la sua rappresentazione nello spazio della concentrazione risulta anch'essa pressoché compatta. In tal caso la diffusione aiuta soltanto la tendenza generale del reagente a rimanere in uno stato stazionario rosso, spazialmente omogeneo.

Argomenti come questi devono e possono essere trattati più accuratamente. Ad esempio, essi non vanno bene se i coefficienti di diffusione di *A* e *B* sono troppo disuguali, perché si formano in tal caso delle instabilità che distruggono la semplicità del ragionamento. Per accertare se un'onda a spirale stabile può davvero evolversi a causa di una cinetica di eccitazione congiunta a una diffusione illimitata occorre soltanto seguire le equazioni nei successivi momenti in ciascuno dei numerosi minuscoli glob della capsula, aggiungendo ogni volta un piccolo termine matematico per la descrizione dello scambio di molecole tra i glob adiacenti in proporzione alla loro differenza di concentrazione. In teoria, ciò appare molto semplice, ma in pratica è indispensabile ricorrere a un elaboratore per effettuare i calcoli e per mettere in grafico i risultati.

Subito dopo che avevo abbozzato i programmi necessari, ebbi la fortuna di incontrare Patrick Murphy, uno studente della Purdue University, dotato di non comuni attitudini a programmare in modo efficiente gli elaboratori digitali. Murphy elaborò le soluzioni numeriche del modello semplice che avevo delineato e i risultati confermarono le nostre aspettative (*si vedano le illustrazioni da pagina 76 a pagina 78*).

Simili onde rotanti sono la conseguenza di una varietà di schemi cinetici più complicati, sempre che gli stessi posseggano essenziali caratteristiche di eccitabilità: le traiettorie di un glob nello spazio di concentrazione devono convergere verso un singolo stato stazionario, non discosto dalla soglia oltre la quale le traiettorie trovano il loro tragitto di ritorno allo stato stazionario soltanto dopo una lunga e velocissima odissea attraverso concentrazioni estreme.

Questa interpretazione mi sembrava ragionevole finché non mi capitò di scoprire un'anomalia in alcune vecchie fotografie di onde nel reagente Z. Era spesso successo che, prima che uno strato sottile di reagente si organizzasse in spirali ad evolvente, le forme d'onda periodiche prevalenti assomigliassero molto ad anelli e a spirali allungati (*si veda l'illustrazione in basso a pagina 72*).

La spirale allungata assomiglia all'ordinaria spirale ad evolvente ad eccezione del fatto che il suo punto terminale interno non è aderente a un perno. Invece di ruotare una volta sola con periodo τ_0, il punto terminale si muove rapidamente avanti e indietro con periodo τ_0 lungo un arco a forma di fessura, della lunghezza di 10 o 20 millimetri circa. Cosa ancor più sorprendente, gli anelli allungati emergono non da un punto *pacemaker*, ma da una fessura che si divide a metà prima verso destra e poi verso sinistra (completando un anello) e poi ancora verso destra e verso sinistra, sempre con periodo τ_0.

Entrambi i tipi di sorgenti allungate si possono spezzare in un numero maggiore di sorgenti di anelli e spirali meno allungati. In casi del genere, la parità, definita come la differenza tra il numero di spirali in senso orario e quello in senso antiorario, risulta conservata. Negli intervalli fra tali frammentazioni, le sorgenti di anelli e spirali allungati si accorciano continuamente verso una maggiore simmetria circolare. La sorgente di spirali si contrae verso il nucleo, producendo un'onda ad evolvente. La sorgente di anelli, tuttavia, svanisce prima che i suoi anelli conseguano la simmetria perfetta. Le onde che erano appena state emesse si propagano dal centro a velocità costante, lasciando un disco rosso che si allarga, là dove un istante prima si stavano formando onde blu a intervalli perfettamente regolari.

Poiché il colore rivela la concentrazione della fenantrolina reagente, sembra del tutto ovvio che queste onde descrivano l'evoluzione di un insieme di isobare di concentrazione. Tuttavia non fui capace di dare un'interpretazione a queste anomalie che potesse andare d'accordo con una cinetica plausibile e con la continuità nello spazio bidimensionale. Che cosa non andava bene? La soluzione del dilemma sembra ovvia in un esame retrospettivo: abbandono del principio « autoevidente » che i contorni di colore riflettono le concentrazioni chimiche. Ciò è vero in un mezzo strettamente bidimensionale, ma nel reagente Z la profondità anche di un millimetro potrebbe essere importante. Nel caso esistesse una struttura spaziale nel senso della profondità, allora il colore visto in proiezione in tale profondità indicherebbe soltanto la concentrazione media dei reagenti in ciascuna colonna verticale di fluido.

Per saggiare questa possibilità, schiacciai una gocciolina di reagente Z tra due lastre parallele di Plexiglas separate da tre minuscoli cilindretti in modo da poter tagliare e rompere le onde, facendo rotolare una lastra sul-

l'altra. Si crearono e gradualmente decaddero sorgenti allungate, ma soltanto in strati più spessi di circa il diametro del nucleo di una spirale. In film più sottili non si osservarono sorgenti, oppure si può dire che esse si decomponevano istantaneamente per formare onde le cui estremità libere incominciavano immediatamente ad arrotolarsi nelle consuete spirali ad evolvente.

Immaginiamo che nuclei di spirali possano stare inclinati in strati di liquido sufficientemente spessi per poterli contenere. Nelle tre dimensioni, allora, il nucleo dovrebbe estendersi attraverso il liquido come un filo. L'onda emergente dal nucleo dovrebbe essere come un gomitolo che si svolge dal filo: la spirale ad evolvente è solo una sezione trasversale dell'onda a rotolo (*si veda l'illustrazione in alto a pagina 81*). Alcuni tentativi fatti dimostrano che, nella sua proiezione attraverso uno strato di liquido due o tre volte più spesso del diametro del nucleo di una spirale, un'onda a rotolo, il cui nucleo giace inclinato tra un'interfaccia e l'altra, dovrebbe assomigliare a una spirale allungata emergente da una sorgente simile a una fessura. Un rotolo il cui nucleo tocca una sola interfaccia di due punti, essendo curvato a U dovrebbe assomigliare a un insieme di anelli allungati attorno a una sorgente a fessura. Tutte queste onde avrebbero il periodo τ_0, della rotazione del rotolo.

Che cosa succede quando il nucleo del rotolo giace esattamente parallelo a un'interfaccia vicina? Se esso è abbastanza vicino, sembra che esso graviti dentro l'interfaccia e che svanisca. Una spiegazione di questo comportamento verrà presentata in seguito; però prima ci occuperemo delle sue implicazioni nei riguardi del decadimento delle sorgenti allungate.

Se il nucleo del rotolo di una sorgente di anelli allungati viene a trovarsi troppo vicino alla interfaccia adiacente, tanto da gravitarvi dentro, rimarrà una coppia di sorgenti di spirali, ciascuna lunga circa la metà di quella originale. Una di esse ruoterà in senso orario e l'altra in senso antiorario. Alla sorgente di anelli viene assegnato il valore 0 per la sua parità, come si disse in precedenza; la spirale che ruota in un senso ha la parità 1 e quella che ruota nel senso opposto ha la parità –1. In questo evento, pertanto, la parità risulta conservata: $0 = 1 - 1$. Se il nucleo di un'onda a spirale allungata (parità = 1) passa troppo vicino a un'interfaccia nel suo cammino verso l'altra interfaccia, si può similmente spezzare in una sorgente più corta di spirali allungate (parità = 1) e in una sorgente di anelli allungati (parità = 0). La parità risulta conservata: $1 = 1 + 0$. Se una sorgente di anelli non molto allungati (parità = 0) gravita entro l'interfaccia che le sue due estremità stanno toccando (il che è possibile se le estremità sono tra loro vicine e la sorgente sta generando anelli pressoché simmetrici) la sorgente di anelli può svanire completamente. La parità risulta conservata: $0 = 0$.

Ma perché mai il nucleo del rotolo dovrebbe essere attirato verso la sua distruzione alle interfacce vicine? Abbiamo visto che una capsula quadrata di un reagente idealizzato, caratterizzato da gradienti di concentrazione di *A* e *B* che si incrociano nella regione ambivalente, ruoterà stabilmente nello spazio di concentrazione per il fatto che le traiettorie assunte dai glob reagenti senza alcuna diffusione divergono abbastanza dalla regione ambivalente da poter controbilanciare l'effetto omogeneizzatore della diffusione. Viceversa, il reagente, i cui gradienti di concentrazione non si sovrappongono alla regione ambivalente dello spazio di concentrazione, decade semplicemente in uno stato omogeneo di riposo con i suoi glob che seguono delle traiettorie tutte dirette pressoché nella stessa direzione. Ovviamente, tra questi due casi estremi, vi sono quelli intermedi nei quali il reagente ruota per breve tempo prima di ritornare nello stato stazionario. Per creare tali basi, la capsula di reagente viene spostata fuori centro nello spazio di concentrazione e inoltre con un lato vicinissimo alla regione ambivalente. In altri termini, il nucleo della spirale viene spostato vicinissimo al lato della capsula. Non appena il reagente viene trasportato attraverso e al di là della regione del nucleo, quest'ultimo viene trasportato verso il lato della capsula. Sembra che, almeno in questo semplice modello bidimensionale, il nucleo della spirale sia attratto verso i contorni vicini del mezzo dove la diffusione agisce asimmetricamente.

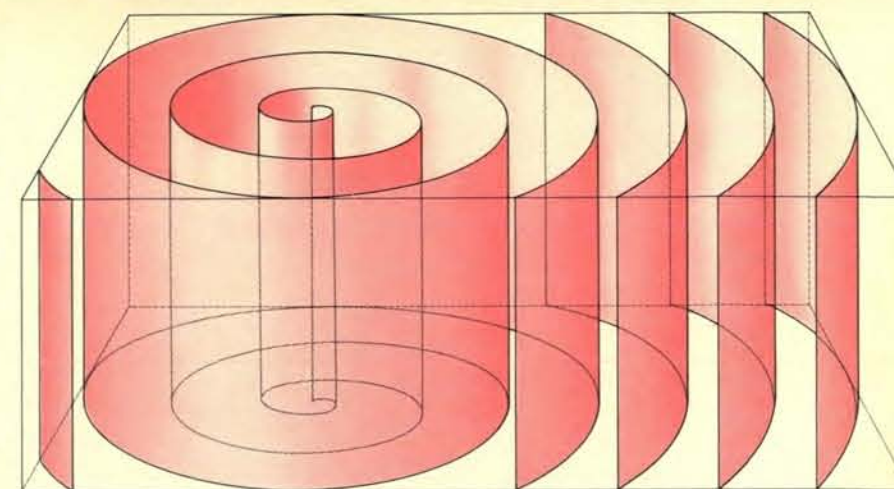
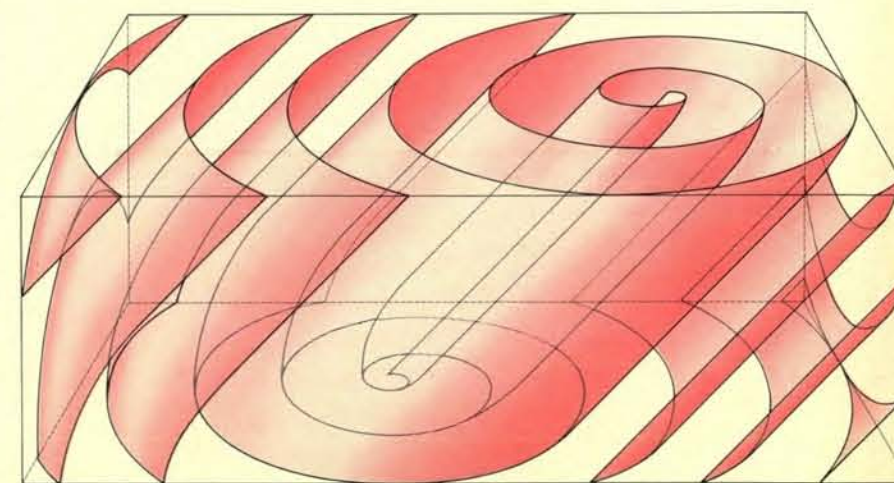
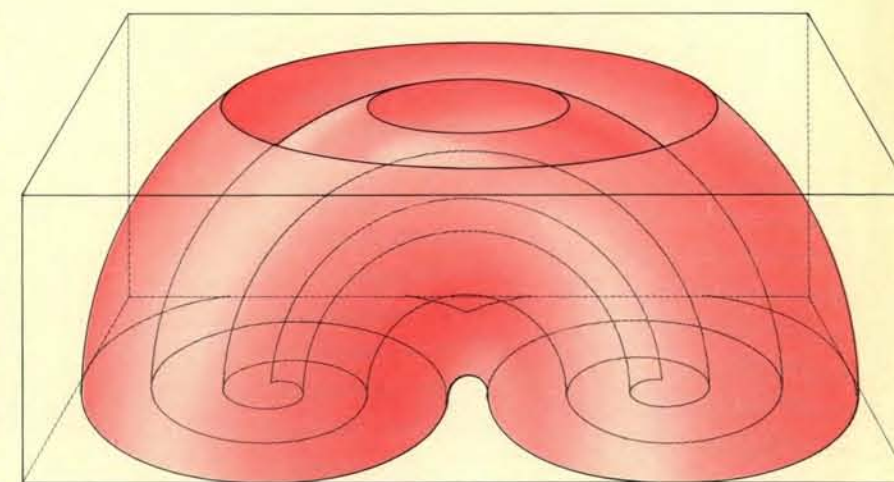
Il nucleo del rotolo vaga attraverso il reagente da una interfaccia all'altra, e poi indietro (*in alto a sinistra*). La linea colorata scura è l'asse del rotolo vista lateralmente, le linee chiare sono le onde emergenti. Se il nucleo si spinge entro la interfaccia superiore (*in alto a destra*), rimangono due sorgenti di spirali allungate, rotanti in senso opposto. Se il nucleo si estende da una interfaccia all'altra (*nel centro a sinistra*), apparendo come una sorgente di spirali allungate, può spezzarsi in una sorgente di spirali più corta e in una sorgente di anelli allungati (*nel centro a destra*). Una sorgente di anelli allungati ad asse cortissimo (*in basso a sinistra*) può gravitare entro l'interfaccia adiacente e svanire (*in basso a destra*), lasciando un'area di soluzione rossa.

Se nello spazio reale tridimensionale il nucleo del rotolo vaga tra interfacce parallele, talvolta accostandosi moltissimo a una di esse, ci si potrebbe aspettare che esso penetri entro l'interfaccia lungo tali aree, frantumando il filo dal quale emerge l'onda a rotolo. Come si è visto, un simile evento potrebbe spiegare le sorgenti allungate e il modo con cui esse decadono. Per accertare se un particolare modello analitico fa effettivamente queste cose, sarebbe necessario integrare numericamente le equazioni per le reazioni e per la diffusione in un reticolo tridimensionale di glob partendo da concentrazioni iniziali opportunamente stabilite. Occorrerebbe, inoltre, un reticolo particolarmente fine, poiché stiamo indagando una instabilità e dobbiamo essere sicuri che ciò non dipenda, invece, dalla nostra suddivisione del reagente in celle discrete.

Un microscopio binoculare può aiutare molto a confermare la propria convinzione in queste argomentazioni alquanto fantasiose. L'osservazione visiva, tuttavia, può raramente estendersi su numerose rotazioni di un'onda a rotolo. Il più leggero disturbo termico, così come l'agitazione meccanica, distruggono il disegno che si sta sviluppando, sempre che ciò non accada prima a causa di una bollicina di anidride carbonica che si libera. Inoltre, anche l'interpretazione del disegno in tre dimensioni, allorché si tenta di mettere a fuoco il microscopio su campi mobili di colore trasparente, concede molto all'immaginazione. Non si potrebbero forse fissare in qualche modo queste onde e poi tagliarle in sottili sezioni per una ricostruzione in serie, nel vecchio, onorato stile dell'embriologia classica?

Il reagente può venire gelificato mediante biossido di silicio colloidale senza alterare molto i processi di reazione o di diffusione; se fosse allora possibile arrestare completamente la reazione a un momento interessante, il taglio del reagente con un microtomo congelante compirebbe il resto. Trovai che anche l'azoto liquido raffreddava il gel troppo lentamente per poterne prevenire la cristallizzazione che, separando i vari composti chimici, distrugge il disegno.

Il reagente che si era versato sul banco di laboratorio sembrò mantenere ragionevolmente bene le onde tra le fibre della carta della mia agenda. Un leggero miglioramento lo ottenni con un gran numero di fogli di carta da giornale, di carta per sigarette e di carte da filtro. Infine, sperimentai i filtri Millipore, composti di esteri di cellulosa chimicamente inerti, crivellati da trafori interconnessi di diametro inferiore al micron. Essi assorbirono il liquido reagente e misero in evidenza onde così nette e colorate come mai erano state viste. Il legame meccanico del fluido nei minuscoli trafori impedisce al liquido di muoversi velocemente. Le onde si propagano, con fronti ripidi, a velocità invariabile, anche se il filtro viene maneggiato liberamente. Sovrapponendo i filtri l'uno sull'altro, aderiscono strettamente tra loro e conducono le onde nelle tre dimensioni proprio come il liquido libero.

A quel tempo John De Simone, Da-

L'onda a rotolo emergente da un asse del rotolo verticale apparirebbe come una spirale ad evolvente se si osserva la soluzione dall'alto. Ogni superficie colorata rappresenta il repentino fronte d'onda dal rosso al blu, mentre si propaga attraverso il reagente.

La spirale allungata potrebbe essere un'onda a rotolo emergente da un asse del rotolo inclinato. Visto dall'alto, il nucleo tenderebbe a confondersi in una sorgente a fessura.

L'asse del rotolo a forma di U tenderebbe a emettere onde a rotolo che, dall'alto, apparirebbero come onde concentriche emergenti da una sorgente a fessura che si divide prima a sinistra e poi a destra, mentre il nucleo ruota sul suo asse. Se i due bracci della U si avvicinassero gradualmente tra loro e la U diventasse più piccola, le onde apparirebbero sempre più circolari. Infine, se la U gravitasse, nella sua contrazione, entro la stessa interfaccia, la sorgente svanirebbe repentinamente, proprio un istante prima che le onde emesse dalla stessa potessero assumere una perfetta simmetria.

vid L. Beil e L.E. Scriven dell'Università del Minnesota trovarono il modo di fissare chimicamente le onde del reagente Z in membrane di collodio. Sebbene non mi fosse possibile usare il loro metodo nei filtri Millipore, il fatto stesso che detto metodo esistesse mi diede coraggio. Emersero due procedimenti di fissaggio. Secondo uno dei processi, dell'acqua fredda satura di sale arresta la reazione e precipita la fenantrolina ferrosa rossa nel filtro, mentre tutto il rimanente si scioglie. Il filtro viene allora fatto galleggiare su una soluzione diluita di ioduro in un recipiente contenente iodio. Non appena la fenantrolina ferrosa si ridiscioglie, essa si diffonde sulla superficie e si combina fisicamente con lo iodio precipitando sulla superficie stessa sotto forma di un film dorato di aspetto metallico. L'altro procedimento è più semplice, più veloce e di maggiore affidamento, sebbene meno colorato: la reazione viene arrestata e la fenantrolina ferrosa viene fissata in una soluzione a freddo al 3 per cento di acido perclorico, mentre tutto il resto si dissolve. I filtri essiccati vengono poi schiariti con olio.

Questa tecnica ha le sue limitazioni, quali la difficoltà di poter manipolare delle pile di filtri poco più alte di un millimetro, e il fatto che bollicine invisibili di anidride carbonica cominciano prontamente a separare i filtri in molte zone. Inoltre, se si apre la pila, l'esperimento termina per il fatto che si rompe la connessione tra le onde che attraversano i filtri contigui. Il rotolo più facile da introdurre in un mezzo stratificato orizzontalmente è un anello a rotolo orizzontale: una onda a rotolo il cui lungo asse si chiude ad anello attorno a un asse verticale. L'anello a rotolo è l'onda a rotolo a forma di U unita alla sua immagine speculare (*si veda l'illustrazione in questa pagina*).

Che probabilità c'è che un sottile filo di un'unica organizzazione cinetica vaghi attraverso una soluzione reagente e che esattamente si congiunga con l'altra sua estremità, formando un anello? L'intuizione dice che non è molto probabile. Quando il fenomeno viene esaminato in termini chimici, tuttavia, sembra del tutto naturale. Consideriamo un volume di soluzione nel quale la concentrazione di *A* vari continuamente da un luogo all'altro in modo complicato, quasi causale. Il liquido risulta stratificato dai suoi livelli di concentrazione di *A* in una serie di lamine attorcigliate, arricciate, ciascuna delle quali è isobara bidimensionale di concentrazione fissa. Lo stesso accade per *B*: le isobare di concentrazione di *B* dividono il liquido in altre serie compenetranti di superfici curve. Là dove entrambe le concentrazioni stanno simultaneamente passando attraverso la regione ambivalente, si ha un pezzetto di nucleo di rotolo. Il nucleo del rotolo segue l'intersezione di quelle isobare critiche di *A* e *B*. Inoltre, se due superfici curve s'intersecano del tutto nello spazio tridimensionale, esse lo fanno tipicamente lungo anelli chiusi, ovvero frammenti di anelli chiusi interrotti dai confini del mezzo. Sembra che se gradienti di concentrazione ragionevolmente semplici vengono imposti a un volume di reagente Z, la configurazione tipica di un qualsiasi nucleo di rotolo risultante dovrebbe essere una serie di lunghi fili, molti dei quali chiusi ad anello. Le illustrazioni di pagina 81 possono allora essere viste come frammenti di anelli a rotolo.

L'anello a rotolo orizzontale (*a*) con asse del rotolo circolare (*linea continua in colore*) e asse della cavità (*linea tratteggiata in colore*) attraverso la cavità «a ciambella» può essere provocato in una pila di filtri Millipore imbevuti di reagente Z. Tagliando con un piano verticale (*b*) la pila di filtri (*area grigia*), l'anello a rotolo apparirebbe come due spirali gemelle ad evolvente. Se i filtri fossero tagliati nel piano orizzontale (*c*), l'anello a rotolo apparirebbe come una serie di anelli concentrici, i cui fronti d'onda si propagherebbero nelle direzioni delle frecce.

Immaginiamo una sezione orizzontale di un anello a rotolo; qui tutte le onde sono dei cerchi e quelle più interne si propagano in dentro. Queste onde che si propagano verso l'interno risulteranno quando a ciascun intervallo τ_0 un giro del rotolo colpisce di fianco l'interfaccia suddividendosi in due cerchi, che si propagano l'uno verso l'interno e l'altro verso l'esterno. Tali strutture si formano spesso nel reagente Z, quando lo si agita dolcemente, ma risultano piccole e circondate da onde più complicate. C'è un sistema per poter osservare anelli a rotolo isolati e abbastanza grandi?

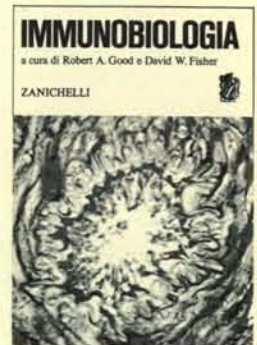
È possibile effettuare questo esperimento con filtri Millipore imbevuti di reagente Z. Si può provocare un'onda cilindrica in una pila di cinque filtri, toccando il centro della pila con un filamento caldo. Allorché l'onda cilindrica risultante (che appare circolare sul filtro superiore) si è propagata sino al diametro desiderato, si aggiunge sulla sommità della pila una seconda fila di cinque filtri, tutti ancora omogeneamente allo stato stazionario rosso. Là, dove l'onda cilindrica confina con il reagente uniformemente rosso, viene suscitato il giusto tipo di gradiente di concentrazione incrociato. Poiché tutto l'esperimento è simmetrico rispetto all'asse verticale della pila, il nucleo del rotolo (se è stabile) deve estendersi in un cerchio orizzontale. Dopo un breve tempo pronosticabile, e poi a intervalli τ_0, si vede un'onda circolare alla sommità del filtro, pressoché direttamente sopra il posto in cui l'onda cilindrica si trovava quando le due pile sono state unite. Quindi l'onda si spacca in un cerchio che si propaga verso l'interno e in un cerchio che si propaga verso l'esterno. La periodicità dell'evento rivela la rotazione di un nucleo intorno alla sorgente di queste onde.

L'assenza di estremità libere e di spirali ad evolvente sulle superfici visibili, superiore e inferiore, dei filtri, indica che il nucleo del rotolo deve rimanere nella pila e che perciò deve chiudersi su se stesso ad anello.

Se questa pila contiene adesso un anello a rotolo orizzontale, che cosa ci si può aspettare di vedere aprendo lateralmente i filtri Millipore? Due caratteristiche dovrebbero essere preminenti. La prima, che gli anelli interni ed esterni dovrebbero apparire o svanire repentinamente a coppie, dovunque il fronte d'onda risulti orizzontale e si propaghi verticalmente. La seconda, che dovrebbe esserci un unico anello spaiato il quale svanisce completamente in qualche luogo all'interno della pila, là dove il bordo interno dell'onda si accosta al nucleo del rotolo. Entrambe dette caratteristiche si verificano nelle pile di filtri Millipore.

Tali anelli a rotolo, nei quali il nucleo vaga per un giro di 360 gradi e si congiunge con se stesso senza mai torcersi attorno al suo lungo asse, o senza fare nodi, potrebbero costituire soltanto la più semplice e più facilmente riconoscibile varietà di forme chiuse a rotolo.

Come il lettore dovrebbe già sin d'ora capire, queste forme di colore, minuscole, delicate e rapidamente mobili risultano difficilmente osservabili nei giusti dettagli. La descrizione matematica, inoltre, di mezzi eccitabili non conduce a facili soluzioni. Conseguentemente, la storia che ho qui raccontato non può essere interpretata come la prova rigorosa che uno stato stazionario stabile e omogeneo, da un lato, e anelli a rotolo, dall'altro, costituiscano i modi caratteristici dell'auto-organizzazione in mezzi eccitabili.

Nondimeno, una conclusione sembra emergere chiaramente e cioè che i principi della formazione di disegni nei sistemi chimici e biochimici sono suscettibili di trattamenti sperimentali e teorici secondo una nuova via.

I cacciatori dell'età glaciale in Ucraina

Nel periodo compreso tra 75 000 e 10 000 anni fa, gli uomini cacciavano altri mammiferi nel rigido ambiente dell'Europa orientale e accumulavano ossa di mammut per costruire ripari

di Richard G. Klein

L'uomo moderno ha vissuto per lo più in un'epoca glaciale. Più precisamente, il tipo di essere umano che oggi conosciamo ebbe origine molto tempo fa, nel Pleistocene, che terminò circa 10 000 anni fa. Il Pleistocene è stato un'epoca di spettacolare attività glaciale, ma non ebbe un clima uniformemente freddo per tutta la sua durata. Vi furono diversi episodi glaciali distinti seguiti da intervalli interglaciali: il numero complessivo di queste fasi alterne non è ancora noto con certezza.

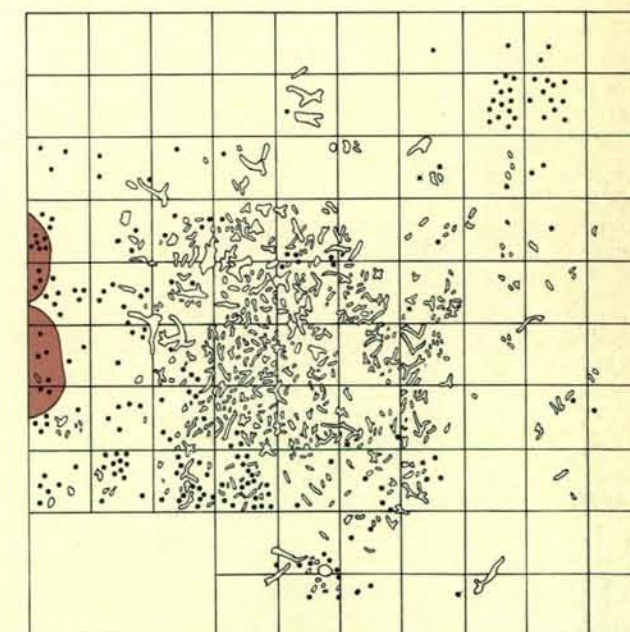
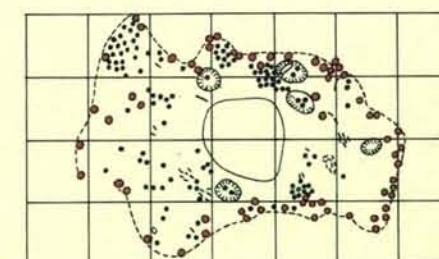
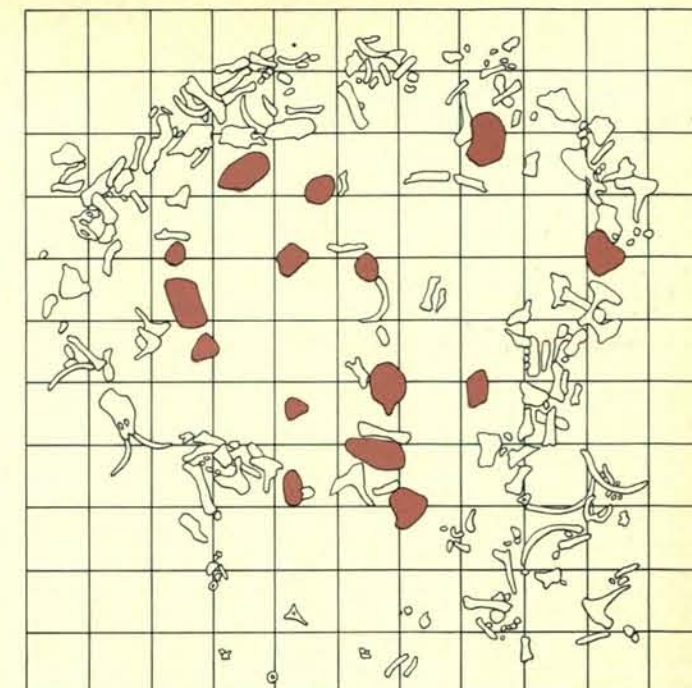
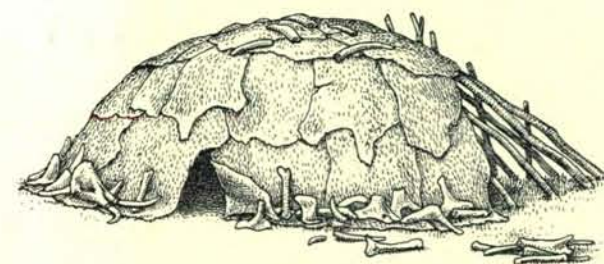
Durante ciascun episodio glaciale, enormi coltri di ghiaccio avanzavano a coprire quasi per intero l'America settentrionale e l'Europa. Quando i ghiacci si ritiravano, il clima degli intervalli interglaciali era simile a quello attuale. La penultima grande avanzata, nota agli studiosi come Riss II, segnò la fine del Pleistocene medio circa 125 000 anni fa. I 115 000 anni successivi appartengono al Pleistocene superiore, un periodo che comprende l'interglaciale Riss-Würm e la glaciazione Würm, in un certo momento della quale, probabilmente tra 45 000 e 35 000 anni fa, sembra abbia fatto la sua comparsa l'uomo moderno (*Homo sapiens sapiens*).

Lo studio sistematico dell'uomo del Pleistocene non ebbe inizio che verso la metà del diciannovesimo secolo, e ciò avvenne in Francia. Oggi, dopo più di un secolo di studi, sia la quantità delle scoperte fatte in Francia, sia l'alta qualità delle ricerche ivi intraprese hanno collocato i dati francesi al primo posto nella maggior parte dei resoconti relativi all'uomo del Pleistocene. Lo studioso viene a conoscenza delle culture progressivamente più recenti quali la Acheuleana, la Mousteriana, la Aurignaziana/Perigordiana, la Solutreana e la Magdaleniana, i cui resti sono stati rinvenuti stratigraficamente sovrapposti in vari siti francesi. Si potrebbe perciò facilmente concludere che il resto del mondo doveva essere ben scarsamente popolato durante il Pleistocene oppure che ben pochi siti di questo periodo glaciale al di fuori della Francia siano stati oggetto di indagine.

In entrambi i casi è vero il contrario. Ad esempio, i primi siti d'occupazione dell'uomo del Pleistocene che vennero scoperti nell'Europa centrale e orientale, molti dei quali spettacolarmente ricchi, furono riportati alla luce quasi nello stesso periodo dei primi ritrovamenti francesi. A causa delle barriere linguistiche, questi siti sono scarsamente noti altrove, ma racchiudono informazioni di vitale importanza per comprendere quanto antichi fossero i tempi in cui l'uomo fu in grado di sopravvivere e forse anche di prosperare nelle condizioni climatiche glaciali dell'Europa. Ciò è forse dimostrato meglio da quanto gli archeologi hanno appreso nell'arco di un secolo da un gruppo di circa 100 siti del Pleistocene situati in Ucraina e nelle regioni adiacenti.

Il più antichi siti d'occupazione del Pleistocene trovati nella Russia europea hanno un'età che si aggira tra gli 80 000 e i 75 000 anni. Essi risalgono all'intervallo interglaciale Riss-Würm (*si veda l'illustrazione a pagina 88*). È possibile che un giorno si scoprano siti ancora più antichi: in Ungheria, Polonia e Cecoslovacchia si sono rinvenute tracce umane vecchie di centinaia di migliaia di anni. Oppure è possibile che solo alla fine del Riss-Würm l'uomo avesse raggiunto la capacità culturale sufficiente a consentirgli di sopravvivere al rigido clima dell'Ucraina. Anche ai nostri giorni gli inverni ucraini sono notevolmente più rigidi di quelli delle confinanti regioni occidentali. Comunque sia, la stragrande maggioranza dei siti d'occupazione dell'età glaciale in Ucraina appartengono alla glaciazione Würm e hanno un'età che si aggira tra i 75 000 e i 10 000 anni. Molti di essi presentano diversi livelli d'occupazione e gli intervalli tra due occupazioni successive sono talvolta lunghi migliaia di anni.

La maggior parte dei siti ucraini si trova nelle principali valli fluviali della regione (*si veda l'illustrazione in*

Tre ripari dell'età glaciale costruiti su un'antica terrazza del fiume Dnestr sono illustrati nella pagina a fronte in pianta (*a destra*) e secondo un'ipotetica ricostruzione. Il primo (*in alto*) venne portato alla luce nel quarto livello del sito noto come Molodova I; 15 focolari (*in colore*) erano circondati da ossa di mammut disposte approssimativamente secondo un ovale. Le analisi del carbone dei focolari, effettuate con il metodo del carbonio-14, rivelano che il quarto livello ha più di 44 000 anni; i manufatti associati sono mousteriani. Secondo la ricostruzione, il riparo era costituito da un'intelaiatura di legno coperta di pelli; le ossa di mammut probabilmente servivano a tenere a posto le pelli. Un secondo riparo (*al centro*) fu riportato alla luce nel terzo livello di un sito adiacente, Molodova V. Un perimetro delimitato da 64 buche (*in colore*) racchiudeva un solo, grande focolare: le analisi del carbone ivi rinvenuto rivelano che il terzo livello ha all'incirca 13 000 anni; i manufatti associati sono del Paleolitico superiore. Secondo la ricostruzione, si trattava di un riparo con un'intelaiatura di legno simile a quella delle tende dei pellerossa; la copertura di pelle era probabilmente assicurata con grossi cavicchi di legno. Il terzo riparo (*in basso*) appartiene al livello superiore successivo dello stesso sito; i due focolari (*in colore*) si trovano su un solo lato dei resti dell'occupazione. La quantità di corna di renna suggerisce che queste ultime dovevano fungere da surrogati delle ossa di mammut nel fissare la copertura di pelli. Le datazioni col carbonio-14 rivelano un'occupazione di 1000 anni successiva a quella del riparo superiore. Le maglie del reticolo sono di un metro quadrato.

questa pagina). Ciò non va attribuito solo al fatto che gli uomini dell'età glaciale frequentassero principalmente le valli fluviali, ma anche al fatto che le valli sono luoghi le cui condizioni favoriscono l'accumulo dei sedimenti e quindi il seppellimento e la preservazione degli antichi siti d'occupazione. L'erosione naturale o le attività della civiltà spesso riportano alla luce i siti sepolti; come accade anche altrove nel mondo, molte delle più importanti scoperte pleistoceniche nell'Ucraina avvennero quando coloro che costruivano strade o scavavano cave d'argilla riportarono in superficie strati contenenti resti animali e manufatti.

Un sito ucraino, che si rivelò costituito da diversi livelli d'occupazione sovrapposti, si trova presso il villaggio di Molodova, sul fiume Dnestr, 250 chilometri a sud-est di Leopoli. Esso fa parte di un gruppo di siti analoghi ed è stato denominato Molodova V. Il più recente dei suoi livelli d'occupazione rimase sepolto oltre 10 000 anni fa; i livelli più antichi, che giacciono tra i sette e gli otto metri sotto la superficie, furono abitati per l'ultima volta oltre 45 000 anni fa.

Le età della maggior parte dei siti ucraini sono state determinate essenzialmente mediante analisi geologica. Ciò è dovuto da un lato all'elevato costo in Russia della datazione con il metodo del carbonio-14, dall'altro al fatto che le determinazioni d'età che superano i 30-40 000 anni sono tecnicamente difficili da stabilire con questo metodo. Tuttavia, a Molodova V una dettagliata analisi geologica condotta da I.K. Ivanova della Commissione sovietica per lo studio del Quaternario (il periodo che, a partire dall'inizio del Pleistocene, arriva fino a oggi) è stata corredata anche da un gran numero di determinazioni col carbonio-14.

Come nella maggior parte degli altri siti ucraini, le tracce di occupazione umana si trovano stratificate all'interno di depositi di sabbia e di *silt*, trasportati in fondo ai declivi dall'azione combinata del dilavamento e della gravità (*si veda l'illustrazione a pagina 89*). In questo caso la base del declivio è costituita da uno strato alluvionale depositatosi nel periodo in cui il Dnestr si trovava a un livello molto superiore all'attuale. Questa antica pianura alluvionale può essere individuata a valle del fiume fino al punto in cui digrada nei depositi marini formatisi quando il livello del Mar Nero era uguale o anche superiore a quello attuale. Ciò, naturalmente, sarebbe accaduto durante il Riss-Würm quando, come oggi, il livello dei mari in tutto il mondo non era abbassato dalla concentrazione di grandi quantità d'acqua nelle calotte di ghiaccio che coprivano i continenti settentrionali.

Come chiarirono le analisi di Ivanova, poiché la pianura alluvionale si è formata durante il Riss-Würm, le sabbie e i silt che la ricoprono sono necessariamente più recenti. Il fatto che i depositi sovrastanti si siano venuti accumulando durante la glaciazione Würm successiva è suggerito da indizi di diverso genere. Innanzitutto, risulta che alcuni di essi si siano formati quando la terra trasudante umidità scivolava a valle su un substrato gelato. Questo processo, noto come soliflussione, è oggi comune nelle zone subartiche e artiche di permafrost, o sottosuolo permanentemente gelato. Il permafrost potrebbe essersi spinto tanto a sud fino all'Ucraina solo durante la glaciazione Würm. In secondo luogo, nei depositi sono presenti conchiglie di gasteropodi che solitamente vivono in ambienti sempre freschi, e le ossa di animali dissotterrate nei livelli d'occupazione umana appartengono a specie subartiche e artiche come la renna e la volpe polare. Inoltre, alcuni dei resti di carbone trovati negli antichi focolari provengono da tipi di conifere che oggi si incontrano solo in ambienti molto freddi. Infine, le determinazioni fatte a Molodova V col carbonio-14, sebbene non concordino affatto, indicano che la più recente delle zone d'occupazione umana risale a oltre 10 000 anni fa. Perciò tutti i livelli d'occupazione sono precedenti alla fine della glaciazione Würm (collocata convenzionalmente intorno all'8000 a. C.). I manufatti e gli altri resti culturali racchiusi nei silt di Molodova V sono inequivocabilmente opera di uomini che vissero durante la glaciazione Würm.

L'analisi di Ivanova del profilo di Molodova V dimostra un altro punto importante: la glaciazione Würm non fu un periodo di temperature uniformemente basse, bensì di fluttuazioni climatiche. In diverse occasioni il movimento di sabbia e silt a valle verso la terrazza del fiume rallentò o si fermò del tutto. Ogni volta che questo accadeva, lo strato superficiale dei sedimenti era soggetto all'azione degli agenti atmosferici che portava alla formazione di un suolo. I periodi in cui rallentava o cessava la deposizione e si aveva la formazione di terra solida rappresentano chiaramente gli intervalli della glaciazione Würm durante i quali il clima diveniva per breve tempo più caldo. Una conferma di questi episodi è fornita dall'assenza dai suoli di Molodova V di quei gasteropodi amanti del freddo che sono stati trovati nelle sabbie e nei silt intatti.

Diversi tipi di dati raccolti sia a Molodova V sia in molti altri siti dell'emisfero settentrionale indicano che la glaciazione Würm fu costituita da tre parti principali. La prima ebbe inizio con l'insorgere delle condizioni glaciali tra 75 000 e 70 000 anni fa e terminò intorno ai 50 000 anni fa. La seconda parte, consistente in un lungo intervallo di clima fluttuante, più mite ma non veramente caldo, durò da 50 000 fino a 30-25 000 anni fa all'incirca. Seguì quindi la terza parte in cui fu raggiunto il culmine del freddo della glaciazione Würm che continuò fino a 10 000 anni fa.

Gran parte degli stessi indizi che hanno portato a individuare le suddivisioni della glaciazione Würm in Ucraina e nelle regioni circostanti rende possibile anche ricostruire l'aspetto del paesaggio. Ad esempio, i granelli di polline sepolti rivelano che la regione era ricoperta da una vegetazione tipica della steppa, caratterizzata da piante che richiedono condizioni di freddo secco. Le comunità di piante

I principali siti paleolitici dell'Ucraina e delle regioni circostanti si trovano nelle valli fluviali come quelle del Dnestr, della regione Dnepr-Desnà e del Don o presso corsi d'acqua minori e affluenti come il Prut, il Bug, il Pripjat', il Sejm, il Sozh e il Sudost'. Poiché molti siti si trovano l'uno accanto all'altro, la loro localizzazione sulla cartina è inevitabilmente imprecisa. Otto di essi (*in colore*) contengono solo manufatti musteriani, 72 (*in nero*) contengono solo manufatti del Paleolitico superiore. A Molodova I, a Molodova V e in qualche altro sito si sono trovate invece tracce di entrambe le culture.

Il paesaggio dell'età glaciale nel periodo di massimo freddo, durante la glaciazione Würm, era costituito da un'ampia cintura di tundra lungo il bordo della calotta di ghiaccio e da una cintura ancora più ampia di steppa periglaciale separata dalla tundra da una zona di steppa con foreste. In alcune regioni, come le valli del Prut e del Dnestr, crescevano foreste di conifere. Il Mar Nero e il Mar Caspio erano uniti.

Quadro cronologico, espresso in migliaia di anni fa, dei periodi glaciali del Pleistocene medio e superiore nelle Alpi (*a sinistra*) e dei loro equivalenti nella Russia europea. Fino a 100 000 anni fa ciascuna suddivisione della scala equivale a un intervallo di 100 000 anni; dopo questa data l'intervallo equivale a 10 000 anni. I più antichi utensili di pietra del Mousteriano conosciuti nell'Europa occidentale hanno più di 100 000 anni d'età. Tuttavia il Mousteriano non dovette raggiungere la Russia che 75 000 o 80 000 anni fa. In Russia gli utensili caratteristici del Paleolitico superiore probabilmente sostituirono quelli di tipo mousteriano tra 35 000 e 40 000 anni fa.

allora presenti non hanno precisi corrispondenti contemporanei e i ricercatori russi le hanno chiamate « steppa periglaciale ». Gli studi dei sedimenti rivelano che la steppa periglaciale si estendeva a sud fino al Mar Nero. Insieme ai dati ricavati dal polline, ciò suggerisce che la temperatura media del mese di gennaio che, nell'Ucraina odierna, è ovunque al di sotto del punto di congelamento, era allora di otto o nove gradi centigradi più bassa.

Potrebbe sembrare sorprendente che gli uomini dell'età glaciale avessero trovato ospitale una regione con un clima tanto rigido. Il fatto è che la steppa periglaciale ospitava una fauna ricca e variata. Le estati ucraine, anche se brevi, erano abbastanza calde da favorire un'abbondante vegetazione per il pascolo degli animali. Gli inverni, malgrado fossero lunghi e freddi, erano secchi; la vegetazione estiva non restava sepolta sotto un'alta coltre di neve e gli animali potevano nutrirsi tutto l'anno. Nei siti del Pleistocene sono state rinvenute le ossa di non meno di tredici specie di erbivori grandi e piccoli. A giudicare dalla relativa abbondanza delle loro ossa, le prede preferite dai cacciatori ucraini dovevano essere la renna e il cavallo selvatico a nord e il bisonte della steppa a sud.

Gli altri erbivori cacciati come selvaggina comprendevano il rinoceronte lanoso (*Opsiceros antiquitatis*), l'uro (*Bos primigenius*), il bue muschiato (*Ovibos moschatus*), la saiga (*Saiga tatarica*), il cervo nobile (*Cervus elaphus*), il capriolo (*Capreolus capreolus*), il cervo « gigante » (*Megaloceros*), l'alce e sia la capra che la pecora selvatica. Raramente tuttavia si sono trovate in gran quantità le ossa di qualcuno di questi animali. Le ossa di carnivori sono anche più rare, solo quelle della volpe polare e del lupo sono presenti in modo significativo. I carnivori più grossi e pericolosi del Pleistocene, l'orso bruno e il leone, praticamente non compaiono.

Meno comuni ancora sono i resti di uccelli e pesci. È probabile che gli abitanti dell'Ucraina dell'età glaciale non conoscessero le tecniche per accostare questi animali, ma possono esserci altre spiegazioni. Una spiegazione interessante è quella che le valli dei fiumi, dove si trovano praticamente tutti i siti noti, fossero abitate soprattutto d'inverno, quando i fiumi ghiacciati ostacolavano la pesca e gli uccelli più pregiati – gli uccelli acquatici migratori – erano volati a sud.

Probabilmente d'estate i cacciatori si spostavano sugli altopiani tra le valli. Queste aree erano zone di pascolo per i grandi branchi di selvaggina, alimento vitale (sotto forma di carne essiccata o affumicata in grado di conservarsi fino all'inverno) per gli abitanti dell'Ucraina. La ragione per cui non si è trovata traccia di questi insediamenti degli altopiani è forse da attribuire alla loro transitorietà che avrebbe lasciato ben poco di tangibile per l'archeologia o al fatto che solo pochissimi sarebbero stati portati alla luce in seguito alla costruzione di strade o allo sfruttamento di cave.

Le ossa di mammut costituiscono in pratica la caratteristica dei siti dell'età glaciale in Ucraina. In più di un caso fu proprio il disseppellimento di questi resti così tipici e imponenti che fece comprendere agli ignari sterratori di aver scoperto un sito dell'età glaciale. Le ossa di mammut sono molto numerose. Ad esempio, quelle riportate alla luce in otto siti del bacino Dnepr-Desnà appartengono ad almeno 500 animali diversi. Nessun sito ucraino però rivela chiaramente se i mammut venivano uccisi proprio lì. Anzi, vi sono prove del contrario: le analisi chimiche delle ossa di mammut trovate in un sito del Dnepr indicano che questi animali vissero e morirono in millenni diversi. Inoltre molte delle stesse ossa, unitamente a quelle di un altro sito del Dnepr, erano state rosicchiate da carnivori. Il significato è chiaro. Gli uomini ucraini dell'età glaciale probabilmente da un lato raccoglievano le nude ossa dei mammut morti da lungo tempo, dall'altro cacciavano e uccidevano loro stessi gli enormi animali. Ma perché le ossa di mammut erano tanto ricercate?

La risposta è chiara e sorprendente. Praticamente in ogni sito dell'Ucraina dove le ossa di mammut sono comuni, per lo più queste sono state rinvenute disposte in modo da far pensare si tratti di rovine di ripari, molti dei quali piuttosto ambiziosi. L'utilizzazione di ossa di mammut come materiale da costruzione è rivelata anche dalla presenza di quantità sproporzionate di alcune di queste: crani, zanne, mandibole, bacini, scapole e arti.

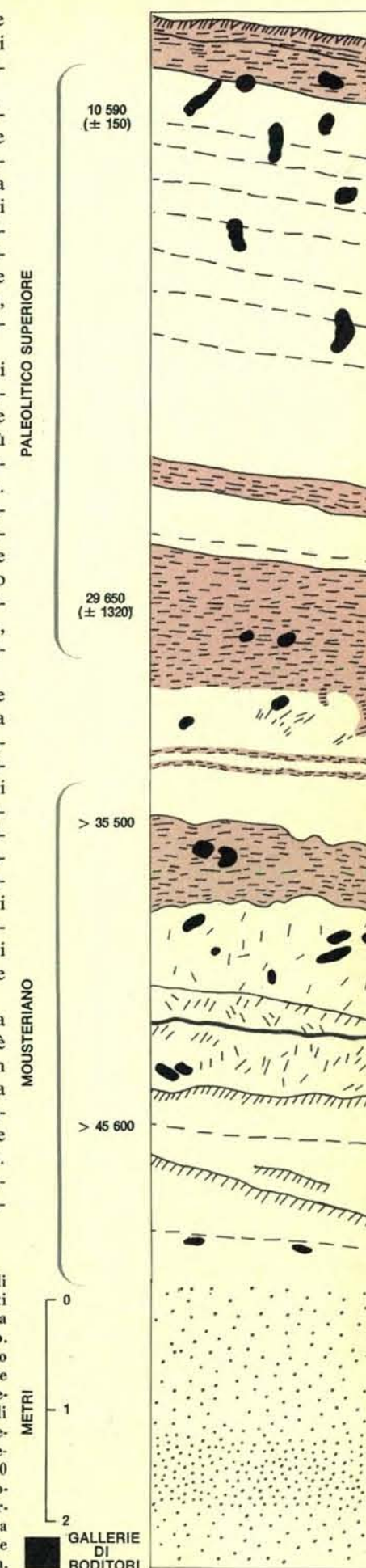
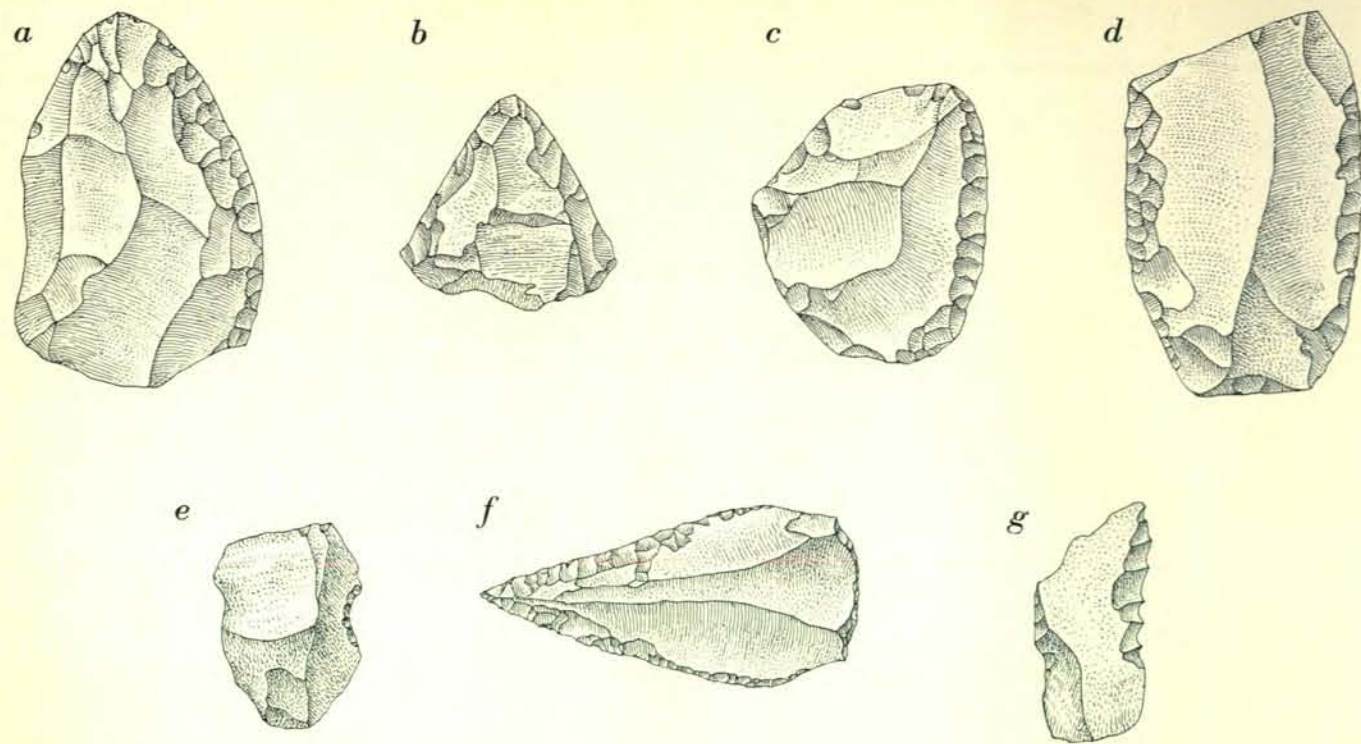
Vi è un indizio proveniente dall'esame di ossa di animali che indicherebbe l'adattabilità dei cacciatori dell'età glaciale alle condizioni invernali. Esso si trova tra le ossa del lupo, della volpe polare e della lepre. Gli scheletri di questi animali sono stati spesso rinvenuti intatti a eccezione delle estremità, completamente assenti. D'altro canto in altri punti del sito sono stati trovati proprio gruppi di queste particolari ossa. Per analogia con le abitudini attuali, si direbbe che gli animali siano stati scuoiati per sfruttare il calore delle pelli.

In conclusione, allora, i resti animali dei siti ucraini dell'età glaciale sembra siano stati accumulati da cacciatori del Pleistocene, molto abili, la cui dieta era ricca di proteine e di grassi, che sapevano apprezzare le pellicce e costruivano solidi ripari. I ripari erano anche riscaldati a giudicare dalle quantità di carbone rinvenuto, indice di focolari, insieme a frammenti d'osso bruciato.

Chi erano i cacciatori? I manufatti trovati nei siti ucraini della glaciazione Würm sono stati attribuiti a due successive e distinte culture. La più antica delle due è nota come Mousteriano e a volte come Paleolitico medio. La più recente è nota invece come Paleolitico superiore. Queste culture successive sono tipiche della glaciazione Würm in tutta l'Europa; il Paleolitico superiore succede al Mousteriano all'incirca tra 45 000 e 35 000 anni fa, un intervallo che tien conto delle diversità delle date da luogo a luogo.

L'utensile mousteriano più comune è il raschiatoio laterale (la cui vera funzione, naturalmente, resta solo oggetto di supposizione). Questi utensili erano costituiti da grandi, irregolari schegge di pietra staccate da un nucleo, solitamente di selce, e quindi « ritoccate » lungo uno o più lati dai quali venivano a loro volta asportate schegge più piccole. Le differenze tra i vari tipi di raschiatoi laterali sono state individuate proprio sulla base di elementi come il numero e la forma dei lati e la qualità e la posizione del ritocco.

Il secondo utensile che si incontra più frequentemente nel Mousteriano è una scheggia con un lato modificato in modo da presentare, sullo stesso, una oppure più tacche. Queste due variazioni sono note rispettivamente come utensili a tacca e utensili dentellati. Si pensa che fossero usati per piallare, per strappare brandelli e per sega-

Questa sezione stratigrafica trasversale di Molodova V (*a destra*) rivela 19 orizzonti culturali all'interno degli strati di sabbia e silt accumulatisi alla base di un pendio. Gli otto orizzonti inferiori contenevano manufatti mousteriani: le analisi effettuate con il metodo del carbonio-14 ne collocherebbe l'età tra più di 35 000 anni e più di 45 000. I manufatti del Paleolitico superiore, provenienti dagli orizzonti più elevati, avrebbero un'età compresa tra 10 500 anni fino a oltre 30 000 anni. Le aree in colore indicano il suolo che si è venuto formando durante gli intervalli più caldi; la zona puntinata rappresenta il materiale alluvionale dell'interglaciale Riss-Würm.

Fra i più tipici utensili mousteriani si annoverano raschiatoi laterali (*a-d*), utensili a tacca (*e*), punte (*f*) e utensili dentellati (*g*). L'uso effettivo di questi manufatti di pietra è sconosciuto, tuttavia è probabile che gli utensili a tacca e quelli dentellati servissero per lavorare il legno. Gli oggetti qui raffigurati provengono da siti mousteriani dell'Europa occidentale.

Gli utensili del Paleolitico superiore, caratteristici per essere scheggiati nella forma desiderata a partire da una stretta lama di selce, differiscono per forma e tipologia dagli utensili mousteriani. Sono qui illustrati alcuni raschiatoi laterali (*a-c*), un bulino (*d*), un punteruolo (*e*), una lama a dorso smussato (*f*), una lunga punta a forma di foglia (*i*) e due punte a intacco (*g, h*). Tutti questi oggetti provengono da siti russi; *h* e *i* provengono dal più basso degli orizzonti del Paleolitico superiore di Molodova V. Come nella figura sopra, tutti i manufatti sono stati riprodotti a due terzi delle loro dimensioni reali.

re. In molti siti mousteriani si sono anche rinvenute punte, rozzamente forgiate a forma di foglia, che spesso hanno una o entrambe le facce rifinite in modo più o meno completo (*si veda l'illustrazione in alto nella pagina a fianco*).

Se si esaminano nel loro insieme gli utensili di pietra del Mousteriano provenienti da diversi siti ucraini, si scopre che fra questi sono presenti alcuni tipi, ad esempio raschiatoi laterali di forma particolare, che non trovano riscontro altrove in Europa. Ma è vero anche il contrario: manufatti come le piccole asce con l'immanicatura trovate in alcuni giacimenti francesi sono sconosciuti in Ucraina. Tali differenze suggeriscono che il Mousteriano non fosse costituito da una singola unità culturale, ma piuttosto da un complesso di sottoculture che si svilupparono in tempi e luoghi diversi.

Una realizzazione del Mousteriano che appare con maggior evidenza nella Russia europea che altrove nel mondo è quella implicita nella raccolta di ossa di mammut: la costruzione di vere e proprie dimore riscaldate. Ad A.P. Chernysh è dovuto lo scavo di un esempio particolarmente significativo di una di queste dimore mousteriane a Molodova I, un sito presso il fiume Dnestr adiacente a Molodova V (*si veda l'illustrazione a pagina 85*). I resti consistono di un ovale di ossa di mammut che racchiude la considerevole area di circa 50 metri quadrati. All'interno dell'ovale Chernysh e i suoi colleghi scoprirono 15 diversi focolari, centinaia di frammenti di ossa di animali, circa 29 000 schegge di selce e una macchia di pigmento rosso. Secondo la ricostruzione di Chernysh, la dimora doveva essere formata da pelli di animali tese su un'intelaiatura di legno; le ossa di mammut servivano da ancoraggio, assicurando le estremità delle pelli al terreno, e probabilmente contribuivano a fissarle anche in altri punti.

Altrove, in Europa e nell'Asia occidentale, le genti che avevano forgiato gli utensili mousteriani seppellivano i loro morti, almeno occasionalmente. Nessuna sepoltura del genere è stata invece rinvenuta in Ucraina, malgrado se ne conoscano alcune in aree immediatamente adiacenti come la Crimea. In Europa, i resti umani associati a manufatti mousteriani appartengono tutti senza eccezioni alla sottospecie umana ancora primitiva dal punto di vista anatomico nota come uomo di Neanderthal (*Homo sapiens neanderthaliaensis*). Ne consegue che generalmente si attribuiscono agli uomini di Nean-

Utensili d'osso sono stati rinvenuti in siti ucraini del Paleolitico superiore, ma raramente in siti mousteriani di questa regione o di altre parti d'Europa. Sono qui raffigurati una punta di proiettile (*a*), due aghi (*b, c*), un raddrizzatore di frecce (*d*), una testa d'arpione (*e*), un punteruolo (*f*) e un martello fatto con corna di renna (*g*). La testa d'arpione proviene da un orizzonte vicino alla superficie di Molodova V, il raddrizzatore di frecce da un orizzonte più antico (Paleolitico superiore) dello stesso sito.

derthal gli utensili del Mousteriano.

I giacimenti di utensili del Paleolitico superiore trovati nei siti ucraini, come i loro corrispondenti in altri luoghi, differiscono notevolmente da quelli mousteriani. Per forgiare i loro utensili, i cacciatori del Paleolitico superiore preferivano schegge lunghe almeno due volte tanto la loro larghezza. Gli archeologi le chiamano «lame». Invece che comprendere soprattutto raschiatoi laterali, utensili a tacca e dentellati, i giacimenti del Paleolitico superiore contengono raschiatoi in punta di lama (lame con una o entrambe le estremità ritoccate), bulini (lame lavorate in modo da presentare un angolo appuntito), lame a dorso smussato e punteruoli (lame lavorate in modo da lasciare una sporgenza aguzza centrale). I giacimenti comprendono anche vari tipi di punte, alcune delle quali con uno o entrambi i lati assottigliati a un'estremità per facilitare la presa, altre ritoccate su entrambe le facce (*si veda l'illustrazione in basso a pagina 90*). Utensili di pietra dello stesso tipo sono stati rinvenuti anche in siti mousteriani, ma sono rari o meno ben fatti o entrambe le cose.

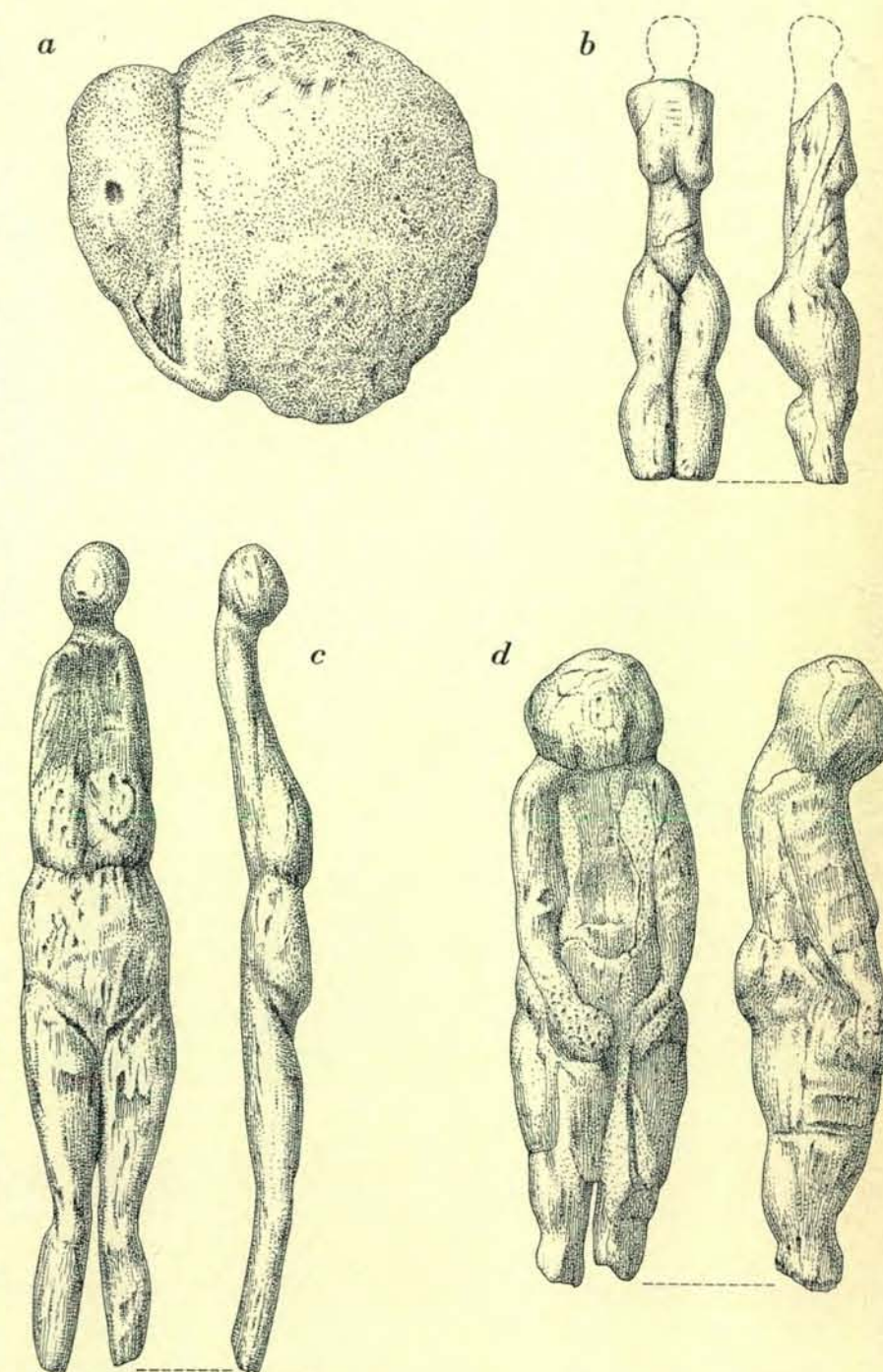
I giacimenti del Paleolitico superiore comprendono anche numerosi manufatti d'osso e corno, praticamente assenti nei siti mousteriani. La forma di tali oggetti fa pensare che si tratti di punte di proiettili, brunitoi, raddrizzatori di frecce, punteruoli e perfino aghi. I cosiddetti aghi sono particolarmente interessanti. Se erano veramente usati per cucire, è altamente probabile che i cacciatori ucraini del Paleolitico superiore portassero indumenti di pelle. Anche al di fuori dell'Ucraina questa conclusione è apparsa di recente praticamente incontestabile. A Sungir', a nord est di Mosca, sono stati portati alla luce tre scheletri di uomini dell'età glaciale avvolti da fili di perline che erano probabilmente cucite sugli abiti che indossavano.

Una caratteristica che i cacciatori ucraini del Paleolitico superiore ebbero in comune con i loro predecessori fu la costruzione di dimore riscaldate formate da un'intelaiatura di legno e appesantite con ossa di mammut. A Molodova V, in uno dei livelli del Paleolitico superiore, sono riconoscibili addirittura le buche che ospitavano i montanti di legno delle strutture. Ma i resti più spettacolari di queste dimore si trovano nei siti del bacino Dnepr--Desnà: qui enormi quantitativi di ossa di mammut sparse su largo raggio circondano o coprono le impronte lasciate dall'uomo sul terreno. Chiaramente in alcuni casi le corna di renna sostituivano le ossa di mammut. I pavimenti delle dimore sono disseminati di buche di varie forme e dimensioni che venivano probabilmente usate come deposito o per seppellirvi i rifiuti.

I resti spettacolari di un riparo del Paleolitico superiore, scoperto a Mezhirich, un sito sul Dnepr a sud di Kiev, comprendevano 385 ossa di mammut che coprivano un'area ovale di circa 3 metri di diametro (*illustrazione in alto a sinistra*). Una volta rimosse le ossa, gli addetti agli scavi portarono alla luce 4600 manufatti di selce, noduli di selce, oggetti d'osso lavorato, frammenti d'ossa carbonizzate, tracce d'ocra rossa, pezzi d'ambra e una buca, profonda 20 centimetri, piena di cenere e carbone. Oltre il perimetro del riparo c'erano altre due aree con focolare (*illustrazione in basso*).

Alcuni indizi dimostrano che altrove in Europa le genti del Paleolitico superiore, al pari di quelle del Mousteriano, spesso seppellivano i loro morti. In Ucraina invece non si sono trovate tombe del genere. Nelle altre parti d'Europa in cui sono state rinvenute delle sepolture, fra cui i siti di aree adiacenti all'Ucraina come Sungir', le ossa in esse contenute appartengono a una sottospecie umana già anatomicamente moderna, l'*Homo sapiens sapiens*. Evidentemente, a un certo punto dell'intervallo di tempo compreso tra 45 000 e 35 000 anni fa, la sottospecie umana più primitiva abbandonò la scena del Vecchio Mondo, per così dire, e salì alla ribalta la sottospecie moderna.

Cosa realmente accadde agli uomini di Neanderthal resta tuttora un mistero. Le testimonianze archeologiche dimostrano che, quanto alle culture materiali rappresentate da utensili e altri manufatti, le manifatture del Mousteriano furono sostituite da quelle del Paleolitico superiore. Ciò significa forse che gli uomini di Neanderthal furono letteralmente eliminati dalla scena dall'uomo moderno? Una conclusione del genere è ritenuta inaccettabile da molti autorevoli studiosi che sostengono che l'uomo di Neanderthal è andato man mano evolvendosi nell'uomo moderno su un ampio fronte nello stesso periodo in cui le manifatture mousteriane si evolvevano in quelle del Paleolitico superiore. Negli ultimi anni, tuttavia, le datazioni ottenute col carbonio-14 per siti come la grotta di Istállóskö, in Ungheria, unitamente alle testimonianze provenienti dai luoghi in cui si era anatomicamente evoluto l'uomo moderno, hanno dato maggior credito all'ipotesi che l'uomo moderno e l'unità culturale del Paleolitico superiore abbiano fatto la loro prima comparsa all'interno di una regione geograficamente limitata: l'Asia sud-occidentale e l'Europa sud-orientale. Secondo tale ipotesi, sia l'uomo moderno sia la cultura materiale del Paleolitico superiore, piuttosto che essersi evoluti in tanti luoghi diversi, si sarebbero irradiati da questo nucleo centrale.

Indipendentemente da come o quando sia iniziato il Paleolitico superiore, prove archeologiche provenienti dall'Ucraina e da altre località russe indicano senza possibilità di dubbio che esso presenta un miglior adattamento del Mousteriano al clima della glaciazione Würm. Questa conclusione trova conferma nella distribuzione dei siti mousteriani e del Paleolitico superiore. L'esplorazione sistematica di quelle regioni settentrionali della Russia dove la glaciazione non ha cancellato tutte le tracce dell'uomo dell'età glaciale non è riuscita a localizzare neppure un sito mousteriano a nord dei 45 gradi

Questi oggetti artistici provenienti da siti russi del Paleolitico superiore comprendono delle «veneri» e un mammut. Il mammut (*a*) e due delle veneri (*c, d*) provengono da Avdeevo, sul Sejm, 400 chilometri a sud di Mosca. La figurina senza testa (*b*) proviene da Eliseevichi, sul Sudost', 350 chilometri a sud-ovest di Mosca. Mentre oggetti simili e ossa incise a motivi geometrici sono frequenti nei siti russi del Paleolitico superiore, né in Russia né altrove in Europa se ne sono trovati in siti del Mousteriano.

di latitudine, mentre siti del Paleolitico superiore sono stati rinvenuti addirittura oltre il circolo polare artico. Il numero molto più grande di siti del Paleolitico superiore, in particolare quelli con « rovine » che suggeriscono un insediamento stabile o semistabile, implica una popolazione complessiva più numerosa e una densità più elevata delle genti che vissero in questo periodo.

A questo punto è necessario considerare un altro tipo di prova. In nessun sito mousteriano, né qui né altrove in Europa, si sono rinvenute opere che si possano con certezza definire artistiche o decorative. Al contrario, siti del Paleolitico superiore, in Ucraina e altrove, contengono numerose figurine d'osso e di pietra che rappresentano esseri umani e animali, unitamente a braccialetti, collane, conchiglie e denti forati e centinaia di oggetti di osso incisi con motivi geometrici (*si veda l'illustrazione a pagina 93*). È possibile che l'assenza di oggetti del genere in una cultura e la loro abbondanza in una successiva sia l'indizio di una variazione biologica, nel caso specifico l'evoluzione verso una maggior capacità intellettuale?

Se si accetta un'ipotesi del genere, la maggior adattabilità del Paleolitico superiore è un esempio di un'interazione retroattiva tra la variazione biologica da un lato e quella materiale e socioculturale dall'altra. Tra gli altri sviluppi è da annoverare quello di nuove strutture sociali sorte probabilmente per consentire l'integrazione di una popolazione più numerosa e più densamente distribuita. Ad esempio le innovazioni sociali possono aver facilitato la cooperazione di gruppo nel cacciare i grandi mammiferi. Con più cibo a disposizione, l'economia venatoria avrebbe sostenuto più persone. Più sottilmente, le nuove strutture sociali possono semplicemente aver incoraggiato una miglior spartizione del cibo tra i vari gruppi in modo da ovviare a temporanee carenze locali. Se i « ricchi » in un certo anno spartivano con i « poveri », l'effetto a lungo termine di questa operazione è alla fine un maggior numero sia di ricchi sia di poveri.

Comunque siano andate le cose, è logico trarre deduzioni demografiche sulla base di un confronto numerico tra i siti del Mousteriano e quelli del Paleolitico superiore. Se si considerano queste deduzioni insieme alle differenze tra le due culture, soprattutto per quanto riguarda le opere d'arte, si può concludere che il Paleolitico superiore rappresenta un passo avanti nell'evoluzione culturale umana.

Vini, vitigni e clima

I vini sono così diversi l'uno dall'altro in primo luogo per le condizioni climatiche e geografiche che caratterizzano le varie zone di coltura e in secondo luogo per la qualità del terreno

di Philip Wagner

Fino a poco tempo fa il vino era per la maggior parte degli americani una bevanda piuttosto esotica. Benché lo Champagne fosse di prammatica ai matrimoni, il vino come componente regolare dell'alimentazione era limitato perlopiù a famiglie agiate, i cui membri avevano viaggiato molto, e a famiglie di origine mediterranea o centro-europea, che avevano conservato le abitudini contratte in patria. Durante l'ultimo decennio la situazione ha cominciato a mutare e gli americani hanno cominciato a scoprire il vino. L'aumento nel consumo di vino è stato diffuso, costante e solo in parte capriccioso e non c'è alcun sintomo di inversione di tendenza.

La generale adozione del vino è così recente che molti di coloro che lo bevono sono ancora incerti quando si tratta di distinguere un vino buono da uno cattivo e un grande vino da un vino semplicemente buono. La curiosità è grande ma la competenza piuttosto rara. Per molti di coloro che si sono convertiti recentemente al vino rimane un mistero il fatto che i bevitori di vino esperti definiscano alcuni vini solo normali e nondimeno continuino a berli con soddisfazione mentre diventano improvvisamente attenti e pieni di rispetto quando ne assaggiano altri.

I giudizi su un vino sono sempre un po' soggettivi, nel senso che quasi tutti hanno i loro vini preferiti. Tali preferenze si fondano però su una base definita e oggettiva, in relazione alla quale diventano possibili le distinzioni. In quanto segue cercherò di fissare gli elementi fondamentali che stanno alla base di alcune di tali distinzioni. Chiunque conosca i princìpi e abbia anche un sistema sensoriale sano e una buona memoria troverà che un vino può significare molto di più di quanto non appaia a prima vista. Le porte sono aperte. Lo stesso accade con la musica, la pittura e la letteratura.

Consideriamo un bevitore esperto che non abbia molto interesse per vini con un retrogusto dolce. È questa la parte soggettiva del suo giudizio. Nonostante la sua preferenza (o pregiudizio), egli è ancora in grado di riconoscere come qualche cosa di eccezionale uno dei grandi Sauternes francesi o dei Trockenbeeven (passiti) tedeschi: un prodotto della congiunzione unica di talune uve maturate in certe condizioni e di metodi speciali e rigorosi di preparazione. Egli riconoscerà anche le differenze esistenti fra un tale vino e un altro «costruito» semplicemente con l'intenzione di farlo assomigliare al primo. Questo è ciò che intendo per distinzione oggettiva.

I vari tipi di uva hanno tutti il nome di genere *Vitis*, il vocabolo latino che indica la pianta della vite. Il genere comprende alcune decine di specie d'uva che crescono in varie parti dell'emisfero boreale. La maggior parte di esse non ha alcuna importanza per la vinificazione, anche se alcune hanno altre proprietà importanti per la viticoltura contemporanea. Il vino che soddisfa gli standard europei viene ottenuto da una sola specie, quella eurasiatica della *Vitis vinifera*, e da pochi ibridi derivati dall'incrocio della *V. vinifera* con altre specie che forniscono frutti simili.

La *V. vinifera* si suddivide in numerosi raggruppamenti, i quali sono spesso così diversi fra loro che molte volte è stato riproposto il problema se si tratti realmente di una specie pura. Se si considera la definizione di una specie come qualcosa di assoluto e non semplicemente come un'espressione puramente convenzionale dettata da ragioni di comodità, allora durante l'evoluzione delle uve da vino ci sono stati sicuramente molti incroci. Per ora è nondimeno sufficiente dire che i gruppi e gli individui abbracciati dal nome *V. vinifera* appartengono alla stessa specie generale e che le loro differenze si spiegano con l'eliminazione da parte della natura degli esemplari più deboli (selezione naturale) e con la scelta da parte dell'uomo dei tipi migliori (selezione artificiale). Considerati insieme, questi due tipi di selezione spiegano l'esistenza di vini così radicalmente diversi fra loro come il Retsina greco e lo Champagne o il Porto e il Beaujolais.

Tutte le varietà di *Vitis vinifera* hanno taluni caratteri comuni. Esse sono decidue. Non sono in grado di resistere a taluni parassiti, e particolarmente alla fillossera e ad alcuni nematodi e sono estremamente sensibili a varie malattie fungine. Per portare a giusta maturazione i frutti hanno bisogno di molto calore e illuminazione solare. La loro resistenza ai rigori dell'inverno è limitata, cosicché la loro sopravvivenza diventa incerta quando l'isoterma di gennaio scende molto al di sotto di un grado centigrado. Le parti del mondo in cui la vite può prosperare e produrre vino sono ristrette, benché l'area complessiva sia molto grande.

Vigneto a terrazze piantato recentemente a Napa, in California, quale appare in una fotografia aerea. La regione ha un clima che è una speciale variante del clima bistagionale del Mediterraneo, con inverni freschi e umidi ed estati asciutte. La Napa Valley beneficia molto della vicinanza alla Baia di San Francisco e dell'influenza di venti freschi del Pacifico che affluiscono attraverso la Golden Gate. Essa trae vantaggio anche dalla sua altitudine ed è la regione della California con la maggior produzione di vino.

La coltivazione di viti della famiglia della *vinifera* nei bacini del Congo e del Rio delle Amazzoni o nel Seskatchewan e nelle steppe siberiane sarebbe altrettanto poco pratica quanto la coltivazione dell'ananas nel Maryland. Il clima ha l'ultima parola.

Per illustrare l'importanza del clima, è opportuno tracciare una linea ondulata con andamento orizzontale da ovest a est su una carta d'Europa (*si veda la figura in queste due pagine*). L'Europa è il continente in cui si produce la maggior quantità d'uve da vino. La linea ha inizio nella Bretagna, a nord della Loira, procede verso est e piega poi bruscamente verso nord-est in prossimità di Chartres includendo Parigi. (A Montmartre c'è un piccolo vigneto.) Procedendo verso est, la linea passa poi in prossimità di Reims e a nord della regione della Champagne, volge leggermente in direzione sud-est verso Nancy in Lorena e poi piega nettamente a nord seguendo il corso della Mosella lungo il confine del Lussemburgo, e poi il tratto, meglio noto, della Mosella tedesca fino a Coblenza.

Successivamente la linea inclina grosso modo verso sud-est presentando per un certo tratto un andamento sinuoso, entra in Cecoslovacchia e poi, giunta sull'Oder, punta bruscamente verso sud, arrivando fin quasi ai sobborghi di Vienna. (Nelle zone al di sopra delle Alpi in cui la vite non cresce la linea è spezzata.) A breve distanza da Vienna c'è un'altra deviazione netta, questa volta verso est. Ora la linea passa a nord della pianura ungherese, supera i Carpazi, attraversa il Dnepr e il Don a una certa distanza dal Mar Nero, passa nella Georgia sovietica a nord della catena del Caucaso, attraversa il Mar Caspio e si perde infine nei deserti del Kazachistan. Questa linea che attraversa l'Europa illustra il limite settentrionale della coltivazione della vite eurasiatica. Esso è determinato dalla concomitanza di due fattori. L'estremo occidentale è soggetto all'influenza del mare, il quale rende più temperato l'inverno ma, a questa latitudine piuttosto alta, mantiene eccessivamente fresca l'estate, cosicché a nord di questa linea la vite non porta a sufficiente maturazione i suoi frutti. Più a est ha il sopravvento l'influenza continentale e quando la nostra linea raggiunge l'Oder il clima invernale è troppo rigido perché la vite possa sopravvivere.

Non si deve credere che il significato di questa linea sia così assoluto che in nessun punto a nord di essa possano crescere viti. Vari incroci con altre specie danno viti i cui frutti maturano più precocemente e che sono in grado di sopravvivere a climi invernali più rigidi; queste viti danno un vino che, se non altro, è accettabile. Viti di questo genere vengono coltivate in località climaticamente privilegiate dell'Inghilterra meridionale e addirittura a Mosca, dove la fiera agricola permanente include un piccolo vigneto. Le viti del vigneto di Mosca, tuttavia, sono ibridi resistentissimi realizzati dal naturalista russo Ivan Mičurin che non hanno però alcuna importanza ai fini della vinificazione e d'inverno devono inoltre essere sepolte per poter sopravvivere. Una situazione assai simile si avrebbe per le parti più fredde della zona temperata dell'America settentrionale. La vite eurasiatica può essere dunque opportunamente coltivata solo in una fascia di varia ampiezza compresa fra la linea del limite settentrionale e una linea formata dal litorale meridionale del Mediterraneo (e, ovviamente, nelle altre parti del mondo che presentano condizioni climatiche simili; in particolare in California, in Cile-Argentina, nell'Africa australe e nell'Australia meridionale).

Fissati questi estremi, vorrei ora tracciare un'altra linea, la quale attraversa anch'essa l'Europa da ovest a est. Questa seconda linea ha un'importanza fondamentale per chi desidera cominciare a orientarsi in questo campo.

La linea divide l'Europa in due aree nettamente distinte dal punto di vista della viticoltura. Essa può essere tracciata con notevole precisione (molto maggiore di quella tentata nella cartina qui riprodotta) benché abbia un andamento assai tortuoso e lungo tutto il suo corso possano essere sollevate discussioni locali a proposito dei punti di riferimento. Stranamente, questa linea non è mai citata nei consueti libri di consultazione nonostante la sua importanza fondamentale nel suddividere i vini europei in due gruppi nettamente distinti.

Questa seconda linea ha inizio in Portogallo a sud di Oporto e segue la valle del Douro, produttrice di porto, risalendola verso est. Una volta superata la parte centrale della penisola iberica, volge verso nord tagliando il distretto spagnolo della Rioja fra Haro e Logroño. Facendo un altro salto, questa volta oltre i Pirenei, riprende in paese basco nella Francia sud-occidentale, puntando a nord-est verso Cahors sul fiume Lot e tagliando la Garonna vicino alla confluenza del suo tributario Tarn. Cahors, che ha legami assai antichi con Bordeaux, più a valle, è un punto di importanza cardinale. Da qui la linea procede a zig-zag superando la regione montuosa del Centre, curva verso nord-est in prossimità

Le regioni produttrici di vino dell'Europa temperata e della regione mediterranea sono illustrate nella presente cartina, nella quale sono state tracciate due linee ad andamento pressappoco orizzontale che corrono da ovest a est. La linea superiore segna il limite settentrionale della coltivazione della vite. La linea inferiore divide l'area in cui la viticoltura risente dell'influenza del clima mediterraneo da quella in cui dominano condizioni di clima temperato. I numeri indicano regioni importanti in cui vengono prodotti vini mediterranei e vini del clima temperato. Gli stessi numeri ricompaiono nella tabella riportata a sinistra della cartina, dove vengono dati i nomi delle località produttrici di vino oppure i nomi dei vini prodotti in queste località.

del corso superiore dell'Allier e della Loira costeggiando il fianco occidentale delle Cevenne e infine arriva a Lione, alla grande ansa in cui la Saône (il fiume del Borgogna e del Beaujolais) si unisce da nord al Rodano. Qui la linea punta verso l'alto seguendo la serpeggiante valle del Rodano fino al lago di Ginevra. Successivamente si arriva in una zona confusa corrispondente alla regione alpina, in cui sarebbe più opportuno usare lineette. I laghi glaciali dell'Italia settentrionale pongono un problema; la linea dovrebbe passare probabilmente a nord di essi ma la questione è discutibile.

Superata la regione dei laghi, il corso della linea ridiventa chiaro. Passando nettamente a nord della pianura veneta, essa entra in Jugoslavia, tagliando fuori la Slovenia settentrionale attorno a Marburgo (Maribor), una parte della quale era un tempo soggetta all'Austria e produceva un vino eccezionale, e muove poi a est lungo il fiume Drava. Seguendo per un tratto la Drava, la linea costeggia il margine meridionale e poi quello orientale della pianura ungherese, supera da un capo all'altro i Carpazi, attraversa la Moldavia e infine si unisce all'altra linea (limite settentrionale) un po' al di sopra del Mar Nero.

Questa linea definisce, dal punto di vista dell'enologo, il limite settentrionale dell'influenza mediterranea. Il bacino mediterraneo ha un clima bistagionale caratterizzato da inverni miti e piovosi ed estati calde, asciutte (praticamente prive di piogge). Ha una flora tipica, esemplificata dall'ulivo, dalla quercia da sughero, dagli agrumi e dai meloni, dall'oleandro e da numerose piante che forniscono frutti squisiti. Esso ha anche un suo proprio gruppo distinto di vitigni, tutti con alberi genealogici che risalgono fino all'alba della storia.

Come ci si potrebbe attendere, il clima mediterraneo e la comune parentela fra i vitigni che crescono in queste regioni conferiscono una marcata somiglianza a tutti i vini mediterranei. Di norma essi sono carezzevoli, morbidi, hanno bassa acidità, un tenore di alcool relativamente elevato e non sono fruttati. Nel loro aroma c'è un sottofondo un po' marcato e terroso ma non sgradevole. La maggior parte di essi ha poco aroma, ma alcuni (del gruppo del moscato) sono molto aromatici. Le differenze fra i vini rossi, diciamo, delle regioni costiere della Spagna, dell'Algeria, della Tunisia, della Linguadoca, della Sicilia, della Dalmazia e della Grecia non sono molto grandi. Il confine settentrionale dei vini di questo tipo è fornito dalla seconda linea che ho tracciato. Il lettore deve di nuovo tenere a mente che questa linea non è assoluta e che pochi chilometri a nord o a sud del suo corso non fanno differenza.

A nord di questa linea cambia tutto e ci troviamo in un mondo diverso. Qui, in una banda stretta e irregolare che si estende da un capo all'altro dell'Europa, la crescita dell'uva è sostenuta da un clima temperato caratterizzato da quattro stagioni nettamente distinte, con inverni freddi, temperature molto variabili durante la stagione della crescita e precipitazioni distribuite su tutt'e quattro le stagioni. Anche i suoli sono diversi perché (pur tenendo conto di differenze anche gran-

CLIMA MEDITERRANEO		CLIMA TEMPERATO	
VITIGNO	ETTARI	VITIGNO	ETTARI
CARIGNAN	182 780	GAMAYS	40 510
ARAMON	158 080	SÉMILLON	29 640
UGNI BLANC	47 425	MALBEC	16 800
GRAND NOIR	39 520	GAMAYS TEINTURIERS	15 810
ALICANTE BOUSCHET	34 580	CABERNET FRANC	13 830
GRENACHE	31 616	FOLLE BLANCHE	11 860
CINSAUT	15 810	SAUVIGNON	10 870
TERRETS	14 820	COLOMBARD	10 870
CLAIRETTES	13 830	CHÉNIN BLANC	10 870
MORRASTEL	10 870	MUSCADET	10 870
MAUZACS	6 420	JURANÇON	9 890
VALDIGUIÉ	4 940	GROLLEAU	9 890
SYRAH	3 260	PINOT	7 900

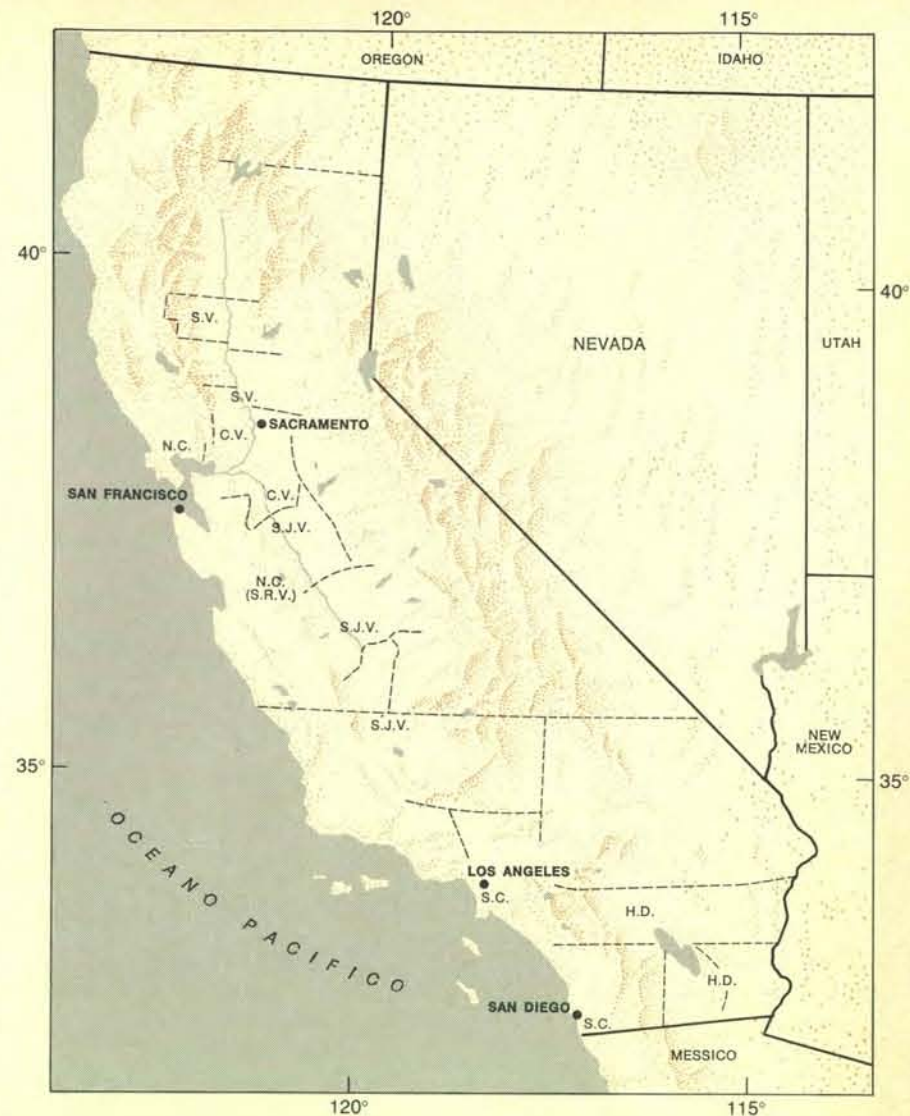
Le località francesi produttrici di vino sono rappresentate su questa carta da puntini neri. La tabella in basso elenca le varietà di vitigni più importanti con l'indicazione del numero di ettari coltivati e del colore del vino: rosso (*in colore*), bianco (*in nero*).

di di origine geologica) i loro caratteri tendono a essere in funzione del clima. Il suolo è modificato radicalmente dal clima. In questa regione anche la vita delle piante è diversa da quella che si riscontra nel Mediterraneo, pur esistendo una parziale sovrapposizione.

Ciò vale anche per l'uva, che ha un aspetto diverso, un diverso sapore e produce vini diversi. I vini sono in generale più fruttati, hanno un tenore alcoolico più basso e una maggiore acidità originale rispetto ai vini mediterranei. Essi hanno anche aromi più sottili e più delicati. Benché il bacino mediterraneo abbia i suoi vini superiori (i Chianti rossi, certi vini del Piemonte e dei laghi italiani, i migliori vini della valle del Rodano e, naturalmente, talune specialità artefatte come lo Sherry e il Marsala), essi sono relativamente pochi. Gli esperti concordano sulla tesi che la maggior parte dei vini veramente fini (i Borgogna, i Chiaretti bordolesi, gli Champagne, e i vini della Germania e dell'Alsazia, per non citarne altri meno familiari, come quelli dell'Austria e dell'Ungheria) sono vini del clima temperato.

La natura generale dei vini prodotti ai due lati di questa linea è determinata dunque da una combinazione fra il clima prevalente e i vitigni a esso adattati e, in misura minore, il suolo. I pregi particolari delle varie qualità di vini all'interno di questi due grandi raggruppamenti sono determinati da microclimi e da microadattamenti delle uve. La regione della Borgogna ci offre una buona illustrazione di questa tesi. I grandi vini borgognoni, rossi e bianchi, sono il prodotto di vitigni coltivati sui pendii sud-orientali e meridionali, esposti al sole e protetti dai venti settentrionali, della Côte d'Or; tali vitigni sono costituiti da alcune sottovarietà di vitigno Pinot nero e Chardonnay. Nella valle della Saône, a poche centinaia di metri verso est, il microclima è diverso, vengono coltivati vitigni diversi e i vini prodotti non sono considerati grandi vini. Lo stesso vale per la parte superiore delle colline e oltre verso occidente.

Contrariamente a quanto si crede di solito, il terreno, benché importante, è nettamente secondario rispetto ai fattori rappresentati dal clima e dal vitigno. Questa relazione risulta evidente dal fatto che numerosi tipi del tutto diversi di suolo si trovano in molti fra i più famosi vigneti del mondo.

L'importanza dello stretto legame fra clima e vitigno può essere illustrata in un altro modo. Se il Pinot nero della Borgogna viene trapiantato in una

ORIGINE MEDITERRANEA		ORIGINE CLIMA TEMPERATO	
VITIGNO	ETTARI	VITIGNO	ETTARI
THOMPSON SENZA SEMI	92 370	COLOMBANO FRANCESE	9 550
CARIGNAN	11 510	CHÉNIN BLANC	5 770
ZINFANDEL	9 515	CABERNET SAUVIGNON	4 595
GRENAGGIO	6 595	PINOT NERO	1 995
BARBERA	5 895	CHARDONNAY	1 830
MOSCATO	5 015	RIESLING BIANCO	1 250
CABERNET RUBINO	4 585	GAMAY BEAUJOLAIS	1 100
PETITE SYRAH	3 295	GAMAY	1 020
ALICANTE BOUSCHET	2 760	SÉMILLON	995
ROSSO RUBINO	2 735	SAUVIGNON BLANC	665
MISSION	2 560	SYLVANER	570
PALOMINO	2 515	GEWÜRTZTRAMINER	465
SALVADOR	1 070	MERLOT	405

Le località produttrici di vino della California sono indicate con iniziali che, dall'alto al basso, stanno rispettivamente per Sacramento Valley, North Coast, Central Valley, San Joaquin Valley, Salinas River Valley, South Coast e Hot Desert. La tabella in basso indica le principali varietà di vitigni distinguendo fra quelle che producono vino rosso (*in colore*) o vino bianco (*in nero*). La California ha un clima bistagionale di tipo mediterraneo, modificato dal fatto che la grande valle interna è quasi chiusa da montagne; questa caratteristica diminuisce le precipitazioni e intensifica il calore.

LOCALITÀ	CLIMA MEDITERRANEO											
	GEN.	FEB.	MAR.	APR.	MAG.	GIU.	LUG.	AGO.	SET.	OTT.	NOV.	DIC.
VALENCIA	**10,4** 30,7	**10,8** 31,1	**13,4** 31,3	**14,5** 33	**18** 28,9	**21,4** 35,4	**24** 8	**24,6** 21,4	**22,6** 49,8	**18** 106,4	**14,2** 57,4	**11,8** 30,9
MARSIGLIA	**5,7** 38,9	**6,4** 41,3	**10,2** 42,6	**12,7** 33,8	**16,9** 42,7	**20,8** 34,3	**23,3** 15,1	**22,6** 30,2	**19,8** 56,5	**14,8** 80,6	**10,2** 59	**7,9** 72,6
NAPOLI	**6,4** 69	**9,9** 64	**10,3** 38	**13,8** 59	**17,4** 59	**21,4** 21	**23,6** 16	**23,8** 27	**21** 92	**16,9** 118	**13** 129	**9,9** 136
ATENE	**10,6** 58,8	**11,5** 25,5	**12,3** 32,1	**16,3** 19,1	**20,9** 29	**25,7** 10,8	**28,6** 2,7	**28,9** 6,2	**24,7** 10,6	**19,8** 68,7	**15,6** 64,9	**12,3** 52,4
SALONICCO	**6,1** 42	**7,5** 29,8	**9,4** 32,8	**14,3** 34,5	**19,6** 44,7	**24,8** 30	**27,4** 18,5	**27,4** 16,9	**22,5** 30,6	**16,5** 62,9	**11,7** 61,2	**8,2** 49,7
ODESSA	**−4,1** 15,7	**−1,2** 21,2	**2** 22,3	**9,4** 24,1	**16,1** 33	**20,2** 38,6	**22,1** 41,6	**21,4** 42,3	**17,4** 10,1	**10,4** 28,6	**4,8** 37,4	**−0,2** 35
SACRAMENTO (CAL.)	**8,4** 112,7	**10,4** 78,1	**12,4** 52,6	**15,5** 48,5	**18,4** 14,9	**21,8** 3,3	**24,4** T	**23,4** 2,7	**22,5** 12,8	**18,1** 13,3	**12,2** 41,7	**8,7** 91,7
BAKERSFIELD (CAL.)	**8,5** 29	**11,1** 29	**13,8** 27	**17,3** 21	**21,3** 6	**25** 2	**29** T	**27,7** T	**24,7** 2	**19,3** 8	**13,3** 12	**9,3** 25
NAPA (CAL.)	**8,6** 123,4	**10,3** 108,4	**11,9** 84,3	**13,6** 43,9	**16** 22,1	**18,3** 5,3	**19,5** 0,25	**19,1** 0,76	**19,2** 4,6	**16,5** 30,9	**12,5** 57,6	**9,4** 124,7
YAKIMA (WASH.)	**−2,6** 30,2	**1,1** 22,1	**5,5** 15,7	**10,3** 11,9	**14,7** 13,7	**18** 20,3	**21,6** 3,3	**21,1** 5	**16,3** 8,9	**10,3** 15,2	**3** 24,4	**−2,7** 28,4

LOCALITÀ	CLIMA TEMPERATO											
	GEN.	FEB.	MAR.	APR.	MAG.	GIU.	LUG.	AGO.	SET.	OTT.	NOV.	DIC.
REIMS	**1,6** 45,7	**3,3** 38,1	**7,7** 48,3	**10** 48,3	**13,3** 53,3	**16,6** 53,3	**18,8** 66	**18,3** 58,4	**15,5** 43,2	**10** 66	**5,5** 58,4	**2,2** 58,4
DIGIONE	**1,5** 61,1	**2,2** 44,8	**7,1** 50	**10,2** 36	**14,6** 54,8	**17,5** 97,6	**19,6** 50,8	**18,6** 82,9	**15,8** 75,9	**10,4** 62,5	**5,6** 63	**3,8** 65,9
NANTES	**4,9** 80,5	**5** 73,7	**8,5** 58,4	**10,5** 47,4	**14** 46,8	**16,9** 50,6	**18,5** 52,8	**18,1** 73	**16,1** 88,1	**12,3** 74,9	**7,9** 86,4	**6,8** 108,5
BORDEAUX	**5,2** 101,3	**5,4** 88,1	**9,5** 67	**11,3** 41,3	**14,8** 63,8	**17,6** 76,6	**19,1** 69,6	**19** 75,3	**17** 87,7	**12,6** 86,3	**8,1** 85,2	**6,8** 125,5
FRANCOFORTE	**−0,3** 53,4	**1,9** 50,6	**6,2** 35,1	**11,3** 39	**14,9** 58,5	**17,7** 76,9	**19,7** 67,7	**19** 74,7	**15,7** 43,9	**10,4** 30,9	**5,1** 62,4	**1,5** 58,8
GINEVRA	**1,1** 63	**2,2** 56	**6,1** 55	**10** 51	**14,1** 67	**17,8** 89	**19,9** 64	**19,1** 94	**15,8** 99	**10,3** 72	**5,7** 83	**2,1** 59
VIENNA	**−0,4** 34,6	**0,1** 34,2	**4,4** 46,6	**9,9** 50,7	**14,4** 55,9	**18** 83,5	**19,9** 87,8	**19,2** 72,4	**15,3** 33,9	**10** 41,6	**4,9** 40,2	**2** 48,7
BUDAPEST	**−0,2** 41,3	**0,6** 38	**5,5** 33,5	**11,4** 39,5	**16,2** 61	**20,2** 94,1	**22** 66,4	**21,3** 56	**17** 39,2	**11,1** 49,8	**5,8** 57	**2,6** 54,4
GENEVA (N.Y.)	**−4,1** 50,2	**−2,8** 63,6	**0,3** 77,6	**8,4** 69,8	**13,4** 71	**19,3** 78,4	**21,8** 78,2	**20,8** 69,9	**16,9** 63,1	**11** 89,2	**4,8** 51,7	**1,6** 60
BALTIMORA	**1,5** 86,3	**2** 73,6	**6,1** 96,5	**12,3** 91,4	**18** 101,6	**22,5** 83,8	**24,8** 106,7	**23,8** 132,1	**20** 83,8	**13,8** 81,2	**7,5** 78,7	**2,1** 76,2
SANDUSKY	**−2,1** 68,9	**−,7** 57,9	**2,5** 73	**10,2** 86,1	**15,6** 79,7	**21,5** 88	**23,8** 87,2	**23,1** 97,6	**19,2** 50,9	**13,1** 52	**5,6** 59,4	**0** 56,1
CHARLESTON (S. CAROLINA)	**9,4** 73,9	**11,1** 88,3	**13,2** 108	**18** 72,7	**22,4** 97,7	**25,3** 130,5	**26,7** 199,8	**26,5** 134,2	**23,9** 164,9	**18,6** 74,5	**13,3** 43,7	**9,7** 75,7
ST. LOUIS	**−0,2** 36,4	**2,5** 64,7	**5,5** 66,9	**13,6** 86,4	**19,1** 83,8	**24,5** 111,6	**26,6** 95,9	**26** 56,5	**21,7** 46,8	**14,8** 63	**6,8** 59.8	**2,1** 42,1
LUBBOCK	**4** 17	**6,2** 15	**9,8** 18	**15,2** 29	**20,1** 81	**25,3** 64	**26,4** 51	**26** 43	**21,9** 60	**16,2** 51	**8,8** 14	**5** 16
ALBUQUERQUE	**1,4** 8,6	**4,8** 10,9	**7,6** 12,2	**13,3** 16,5	**18,4** 20,8	**23,2** 11,7	**25,6** 34,2	**24,7** 29,8	**21,1** 21,8	**14,6** 17,8	**7,1** 5,6	**2,7** 13,2
BISMARCK	**−13,6** 14,3	**−9,6** 12,2	**−4,7** 16,2	**6** 21,5	**13** 60,6	**17,8** 86	**21,8** 48,7	**21,1** 51,5	**14,4** 25,5	**7,9** 15,3	**−2,1** 18,6	**−7,7** 10,7
BOISE	**−0,9** 40,4	**1,4** 32,2	**4,5** 28,6	**9,4** 31,3	**13,9** 43,1	**18,2** 22,9	**23,8** 6,3	**21,9** 6,3	**17,6** 8,5	**11,1** 17,5	**3,6** 26	**−0,1** 33,2
PORTLAND (OREGON)	**4,9** 211,8	**6,6** 118,1	**7,8** 124,7	**11,6** 66,9	**14,8** 52,9	**17,1** 47,9	**20,2** 6,8	**19,7** 22,6	**18,2** 42.2	**13,8** 97,7	**8,3** 142,9	**6,1** 169,5

Le condizioni climatiche dominanti in tipiche località produttrici di vino sono indicate da cifre che danno la temperatura media (*in neretto*) in gradi centigradi e la precipitazione media (*in chiaro*) in millimetri. La *T* che appare talvolta significa tracce.

altra parte della stessa area climatica generale, come la regione della Champagne o l'area attorno a Mâcon o il Médoc o la valle della Loira, il vino risultante potrà avere un certo grado di somiglianza, ma sarà nettamente diverso. Se il vitigno è piantato oltre la linea, nel bacino mediterraneo, la qualità specifica del Pinot nero va perduta e si ottiene un altro vino rosso mediterraneo, non troppo raffinato. Lo stesso fenomeno si produce con altri vitigni famosi del clima temperato, come il Sauvignon, il Riesling, lo Chardonnay e il Cabernet. Inversamente, se vitigni mediterranei come il Grenaggio (o Alicante) e il Carignan (noto anche, in Sardegna, come Cagnulari e nell'Italia centrale e nell'isola d'Elba come Legno duro) vengono trapiantati a nord della suddetta linea, si indeboliscono e non portano neppure a maturazione il frutto.

Questo adattamento dei vitigni giusti a un dato clima o microclima è oggetto di studio dell'ampelografia, la quale attinge innanzitutto alla botanica, e poi alla climatologia, alla geologia e alla storia umana. Consideriamo alcuni fra questi fattori.

Risalendo fino ai documenti fossili, vediamo che tracce di *Vitis* (di specie ora estinte) sono state trovate ovunque in Europa, compresa perfino la Scandinavia. In seguito a un lungo e lento processo di migrazione e di adattamento, altri due gruppi affini di specie si svilupparono nell'Estremo Oriente e nell'America settentrionale. Il legame è evidente poiché tutte le specie del genere *Vitis* tranne una hanno lo stesso numero di cromosomi (38) e sono interfeconde. (I Muscadines del versante sud-orientale degli Stati Uniti sono l'eccezione: essi hanno 40 cromosomi e non si incrociano con le altre specie.)

I gruppi dell'Asia orientale e dell'America settentrionale seguirono un corso divergente e al loro stato puro non presentano alcun interesse per la vinificazione.

L'evoluzione dei vitigni eurasiatici fu interrotta drasticamente dai bruschi mutamenti climatici associati alle glaciazioni. A ogni irrigidimento del clima i vitigni furono costretti a ritirarsi verso sud e sopravvissero solo in due aree di rifugio, una tutt'attorno al bacino mediterraneo e l'altra a sud del Mar Caspio. Ogni volta che i ghiacci si ritiravano verso nord si aveva una nuova espansione di vitigni selvatici, in forme modificate, nelle regioni lasciate libere. Da essi emerse infine una nuova specie (la *Vitis sylvestris*, che viene classificata anche come *V. vini-*

La durata del giorno a varie latitudini nell'emisfero boreale. Le latitudini chiave per il vino sono 30 gradi, che segnano all'incirca il limite meridionale per i vini del Mediterraneo, e 50 gradi, che sono approssimativamente la latitudine di Magonza e di Praga e rappresentano il limite settentrionale per i vini del clima temperato. I vitigni ben adattati della zona attorno ai 30 gradi sono piante che abbisognano di un numero di ore di luce ogni giorno relativamente basso per portare a maturazione i frutti, mentre piante che necessitano di molte ore di luce sono più adattate alle latitudini superiori.

fera ssp. *sylvestris*), la quale si adattò alle molte variazioni climatiche e diede quindi origine a numerosi gruppi distinti. Dai semi di quest'uva che sono stati ritrovati, sappiamo che i frutti di *Vitis sylvestris* venivano raccolti e usati da talune popolazioni preistoriche.

I vitigni selvatici (spesso chiamati lambruschi, dal vocabolo latino *labrusca* che indica la vite selvatica) continuarono a prosperare fino oltre la metà dell'Ottocento e molte viti selvatiche (abròstini) sono diffuse oggi nella maggior parte dell'America settentrionale. Poi essi furono praticamente cancellati dalla fillossera, che era stata introdotta accidentalmente dall'America e contro cui essi non avevano difesa. Oggi ne sopravvivono occasionalmente esemplari isolati in parti dei Balcani e dei Pirenei e in altre località isolate. Trovarne uno è un po' come imbattersi in una tribù neolitica di cui non si avesse finora conoscenza. I vitigni che rimasero nelle aree di rifugio ebbero un'evoluzione del tutto diversa. Erano sensibili ai rigori dell'inverno e, come si addiceva alla latitudine, portavano a maturazione i frutti in giornate con poche ore di luce. Avevano un fogliame liscio, non coriaceo, un fusto massiccio e arborescente e richiedevano un periodo di tempo maggiore per la formazione e la maturazione del frutto.

Accadde allora una cosa drammatica. Al loro stato selvatico le viti sono dioiche, esistono cioè piante maschili e piante femminili e perché si formi il frutto è necessario che la vite femminile sia impollinata da una vite maschile, provvista cioè di stami. Varie migliaia di anni fa, nell'area di rifugio del Caspio, apparvero però uno o più vitigni mutanti che erano monoici: i fiori combinavano il pistillo, od ovario, della vite funzionalmente femminile con gli stami, che portano il polline, del vitigno maschile, cosicché le piante erano in grado di autofecondarsi. (La storia di questo sviluppo è stata largamente ricostruita dall'ampelografo russo A.M. Negrul e dai suoi collaboratori.)

Questa evoluzione fu di un'importanza cruciale per molti aspetti. I vitigni possono essere riprodotti vegetativamente (agamicamente), ossia una talea legnosa metterà radici e il vitigno risultante riprodurrà esattamente il vitigno genitore. Così ogni mutante od ogni ibrido desiderabile può essere moltiplicato in modo uniforme e all'infinito. Fu possibile in tal modo selezionare, coltivare e moltiplicare una varietà monoica preferita e utilizzare tutti i vantaggi dell'uniformità e di un'autoimpollinazione regolare e sicura. È probabile che proprio così abbia avuto inizio la coltivazione della vite.

Un secondo vantaggio dell'ermafroditismo consiste nel fatto che, attraverso la normale impollinazione incrociata, esso può essere introdotto nel corredo genetico di vitigni selvatici vicini che non lo possiedono ma che possono avere altri caratteri desiderabili, ad esempio una buona qualità come uva da vino e uva da tavola. Fu così aperta la strada. Il carattere dell'autoimpollinazione fu diffuso per gradi (ma non in modo uniforme) in tutta la popolazione coesistente di vitigni selvatici dioici. La funzione dell'uomo consistette semplicemente nel selezionare. Una volta che veniva scoperta una nuova varietà desiderabile, la sua riproduzione vegetativa era cosa assai semplice.

La selezione ebbe luogo seguendo preferenze colturalmente determinate. In alcune colture veniva messo l'accento su vitigni che davano uva dai chicchi grossi, preferibili per il consumo diretto o per farli appassire; in altre su vitigni in grado di consentire la produzione di un vino di buona qualità. Negrul e altri hanno dimostrato che le piante caratterizzate da chicchi grandi, tra cui alcune con chicchi privi di semi, erano concentrate a sud del Mar Caspio. Questo gruppo è oggi noto comunemente come *Prol.* (dal latino *proles*, che significa progenie) *orientalis*. I tipi a chicchi piccoli, più adatti per la vinificazione, erano raccolti nel bacino del Mar Nero; il nome assegnato da Negrul a questo gruppo è *Prol. pontica* (dall'antico nome del Mar Nero).

Il percorso evolutivo è stato ricostruito (anche se in modo finora incompleto) principalmente sulla base di studi altamente specialistici sui semi d'uva, i quali sono per l'ampelografia quel che i denti sono per la paleontologia umana, nel senso che si conservano molto bene. I reperti archeologici dimostrano che la raccolta dell'uva selvatica continuò anche dopo l'inizio della coltivazione della vite. Taluni documenti archeologici hanno dimostrato addirittura l'esistenza di un'industria specializzata della vinificazione, che attorno al VI secolo a.C. era diventata l'attività dominante di molti insediamenti, in particolare attorno al Mar Nero. Da località come Chersoneso (cinque chilometri a ovest di Sebastopoli), Myrmekion e Tirithica sono venuti in luce i resti di tutto ciò che poteva servire alla produzione del vino: coltelli e torchi da uva, recipienti in pietra, argilla e ceramica per la fermentazione e l'immagazzinamento e piattaforme per la pigiatura dell'uva con i piedi. Alcuni di questi utensili assomigliano all'attrezzatura usata ancor oggi in talune regioni produttrici di vino e qualsiasi vignaiuolo esperto riconoscerebbe immediatamente tali oggetti.

Da questa patria della viticoltura ebbe origine una graduale migrazione verso occidente, in parte in modo naturale e in parte a opera di persone che portarono con sé talee. Attraverso l'incessante impollinazione incrociata delle viti orientali con quelle selvatiche degli esemplari di *V. sylvestris* incontrati più a occidente lungo la via percorsa, il bacino mediterraneo fu popolato dai vitigni del tipo ancor oggi dominante. La penisola iberica era particolarmente ricca di tali utili ibridi, molti dei quali iniziarono una nuova migrazione in senso inverso lungo il bacino mediterraneo. Tra questi tipi più occidentali (la *Prol. occidentalis*) sono il Grenaggio (Grenache, Garnacho e molti altri sinonimi), il Carignan (Cariñeno e così via) e l'Aramon, che produce oggi fiumi di vino abbastanza comune in molte parti della regione mediterranea.

Un altro capitolo della nostra storia è costituito dall'evoluzione della *V. sylvestris* selvatica nell'Europa temperata in vari raggruppamenti ecologici. I vini della Francia nord-orientale erano diversi da quelli della Francia sud-occidentale e così via. Penetrando nell'Europa temperata, i popoli mediterranei portarono con sé, come d'uso, i loro vitigni. Per la maggior parte le viti non erano adatte al nuovo clima, ma erano in grado di sopravvivere abbastanza a lungo da incrociarsi con i vitigni selvatici indigeni. Dal guazzabuglio che ne seguì derivarono pianticelle che combinavano alcune fra le buone qualità di ciascuna pianta madre: l'autoimpollinazione, la capacità di portare a maturazione i frutti in giornate con molte ore di luce, la capacità di adattarsi alle condizioni del clima temperato, il lussureggiamento caratteristico degli ibridi e un'ampia gamma di nuovi aromi di particolare valore per il vino.

Da queste viti furono selezionati i vitigni coltivati dell'Europa temperata. I nomi e i sinonimi di molti di essi attestano la loro origine selvatica: Plant des bois, Sauvaget, Sauvignon, Savagnin, Sauvagnou, Fer Servadou e Lambrusquet. Uno studioso di etimologia esperto in dialetti europei potrebbe aggiungerne a centinaia.

Un'altra prova di quest'origine casuale si trova nell'immensa variabilità all'interno di quello che i francesi chiamano un *cépage*, il quale non è una varietà singola e uniforme che risalga a una prima vite, bensì un gruppo di individui simili ma non assolutamente identici. Citiamo ancora, per fare un esempio, il Pinot nero. Il nome non si riferisce a un individuo bensì a un gruppo di vitigni diversi aventi un medesimo sfondo genetico. Alcuni di essi sono chiamati Pinot con l'aggiunta di una denominazione di luogo (Pinot d'Aunis, Pinot d'Ambonnay). I nomi di altri aggiungono al nome di luogo l'enumerazione di altri caratteri speciali (Gros Plant dore d'Ay) o il nome di colui che l'ha fatto conoscere (Pinot Renault). Decine di Pinot si trovano nella Champagne, altre decine nella Côte d'Or e non meno di 40 nella sola Svizzera di lingua tedesca. Il totale raggiunge facilmente i mille vini che hanno il diritto di portare il nome di Pinot nero. Lo stesso vale per i nomi di altri famosi *cépages*: gli ubiquitari Chasselas, i Côtes, i numerosi Car-

Questa statua allegorica eretta sul terreno della scuola di viticoltura di Montpellier, in Francia, rappresenta una giovane donna che ne sorregge una più anziana. La giovane rappresenta le radici americane che hanno ridato vitalità ai vigneti francesi, rappresentati dalla donna più vecchia, dopo che, nell'Ottocento, erano stati devastati dalla fillossera. Gustave Fox, uno specialista di viticoltura, dirigeva allora la scuola.

menet (incluso il Cabernet), i Riesling: ogni *cépage* ha una grande quantità di cugini, e di fratelli e di zii, e i singoli individui risalgono a qualche singola vite o mutazione che apparve inosservata o che piacque a qualcuno e divenne popolare. Se tutto ciò sembra molto complicato, la verità è che lo è. Una reale comprensione e un vero apprezzamento dei vini, con le loro differenze vistose e sottili, cominciano solo con la comprensione approfondita di questi infiniti adattamenti del vitigno al clima – e in grado minore al suolo – e degli uomini che lo coltivarono e lo conobbero in modo approfondito una generazione dopo l'altra.

Fino a questo punto ci troviamo di fronte ai meccanismi della selezione naturale, con l'uomo che si accontenta di scegliere fra i risultati che si presentano spontaneamente. La storia ha dovuto affidarsi in vari punti a congetture e inferenze. Passiamo ora a considerare l'intrusione deliberata dell'uomo nel processo genetico nel secolo scorso. Mediante incroci all'interno della specie eurasiatica e ibridazioni fra specie diverse sono state create varietà interamente nuove.

Le devastazioni prodotte nelle viti europee durante il secolo scorso a opera della fillossera segnano il vero punto di partenza. Si manifestò allora il disperato bisogno di trovare rizomi resistenti alla fillossera su cui si potessero innestare le varietà classiche, le quali solo così avrebbero potuto sopravvivere. Ciò significava far ricorso alle viti americane, le quali non venivano ancora selezionate ma erano e sono resistenti in vario grado alla fillossera a causa dell'adattamento a tale insetto autoctono del Nuovo Mondo.

La ricerca di rizomi che siano compatibili e dunque adattabili alle condizioni europee richiedeva un lavoro di combinazione e ricombinazione. La nozione di combinare un carattere, per esempio, della *V. rupestris* appalachiana con un altro carattere, per esempio, della texana *V. Berlandieri* e poi di incrociare la nuova pianta con altre varietà fino alla quinta e alla sesta generazione divenne un luogo comune. Da quest'opera derivarono i rizomi dai nomi criptici (AxR 1, E.M. 41, Rupestris St. George, C. 3309, 5BB e così via) su cui prosperano oggi i produttori di vino d'Europa, California, Australia, Sud-Africa e Cile.

Insieme a questo lavoro si impose uno sforzo divergente e competitivo inteso ad aggirare il bisogno dell'innesto mediante la produzione di ibridi che combinassero la qualità del frutto e del vino dei vitigni europei con la resistenza ai rigori dell'inverno e alle malattie di queste specie straniere. Era prevedibile che tali ibridi non solo estendessero la coltivazione della vite a nuove aree ma anche offrissero vantaggi rispetto ad alcune delle antiche varietà esistenti.

Per entrambi questi motivi il nuovo sforzo incontrò difficoltà. I produttori di vini affermati non desideravano, ovviamente, che nuove aree venissero a far loro concorrenza. In termini di qualità i produttori di grandi vini erano soddisfatti di ciò che già avevano, ora che i loro vitigni erano in grado di sopravvivere dopo l'innesto sulle nuove radici.

A complicare ancor più la situazione, l'ibridazione diede in principio risultati molto mediocri, cosa che ci ricorda l'ammonimento di George Bernard Shaw a Ellen Terry, secondo cui dalla loro unione avrebbe potuto nascere un figlio col cervello di lei e con la bellezza di lui invece che l'inverso, come sarebbe stato più desiderabile. La resistenza al freddo e alla malattia era migliorata, ma a spese della qualità. Il lavoro nondimeno continuò, principalmente a opera di un gruppetto di uomini appassionati.

Alla seconda generazione i risultati di tali incroci migliorarono. Oggi esistono numerosissimi ibridi di questo tipo, spesso dalla genealogia molto complessa, che danno buon vino in condizioni ostili alla *V. vinifera*. Non sono grandi vini: finora nessun ibrido dà un vino paragonabile ai grandi borgognoni e chiaretti e al vino del Reno. Un motivo di questa situazione va visto nel fatto che talune associazioni genetiche trasportano caratteri sia buoni sia cattivi. Bisogna tuttavia ricordare che la massima parte del vino prodotto nel mondo viene ricavato dai vitigni di *vinifera* e che il 90 per cento di esso è ben lontano dall'essere di qualità eccelsa. La tendenza corrente è quella di operare il confronto fra gli ibridi peggiori e i vitigni migliori di *vinifera* e non l'inverso, come sarebbe più giusto.

I produttori europei di vini sono sempre stati implacabilmente ostili a questo sforzo. Nelle regioni europee produttrici di vino è consentito solo di coltivare pochissimi di tali ibridi e un lavoro privato continuo nel campo dell'ibridazione è a tal punto ostacolato da proibizioni e limitazioni da risultare praticamente impossibile. Questo atteggiamento è strano se si pensa ai grandi vantaggi introdotti dall'applicazione della genetica ad altri settori dell'agricoltura e all'allevamento degli animali.

Indubbiamente la viticoltura europea può fare a meno di ibridi. In altre aree, in particolare negli Stati Uniti, l'ibridazione viene praticata attivamente. In California si sta lavorando soprattutto alla produzione di nuovi ibridi all'interno della specie, incrociando vitigni dell'Europa temperata e del bacino mediterraneo nell'intento di ottenere vitigni migliori per la Central Valley, che è fertile ma deserta. Alcuni fra i risultati ottenuti sono stati notevoli. Questo lavoro è stato diretto da Harold P. Olmo, della Università della California a Davis, e da Elmer Snyder del Dipartimento dell'agricoltura degli Stati Uniti. Anche nelle aree della costa occidentale, dove i vitigni eurasiatici si sono adattati in modo superbo, l'esigenza di resistere alla fillossera e ai nematodi ispira un lavoro continuo di ibridazione, interspecifica oltre che intraspecifica.

Altrove, e particolarmente nella stazione sperimentale del Dipartimento dell'agricoltura a Geneva, nello stato di New York, il lavoro nella viticoltura si concentra su incroci fra specie, in cui consiste l'ibridazione in senso proprio. La ragione di ciò consiste nel fatto che nell'America settentrionale la vite eurasiatica non è perfettamente adattata, eccezion fatta per alcune aree, e particolarmente per la California, con le sue molte varianti del clima mediterraneo. Gli estremi climatici di tipo continentale dominanti nella maggior parte del continente sono inospitali per la vite.

Fino all'introduzione dei migliori ibridi franco-americani, con la loro superiore resistenza ai rigori dell'inverno e alla malattia, non fu possibile produrre nell'America settentrionale temperata, col successo che ci si sarebbe potuti attendere, vini paragonabili ai livelli di qualità europei. Oggi si riscontrano successi del genere in molte località americane, in alcuni casi contro ogni verosimiglianza: negli altipiani del Texas, nell'Utah e nell'altopiano dell'Ozark e nella maggior parte dei territori a est del Mississippi, compresi anche il New Hampshire e il Vermont. Non si può essere completamente soddisfatti dei vini che ne risultano perché non esiste ancora alcun ibrido « ideale ». Poiché però i migliori fra gli ibridi esistenti vengono riprodotti e coltivati, poiché la produzione di vino si sta estendendo a nuove regioni e il pubblico accetta ciò che viene prodotto, anche i bevitori di vino americani possono guardare al futuro con ottimismo.

GIOCHI MATEMATICI

di Martin Gardner

Nove impegnativi problemi, alcuni logici, altri no

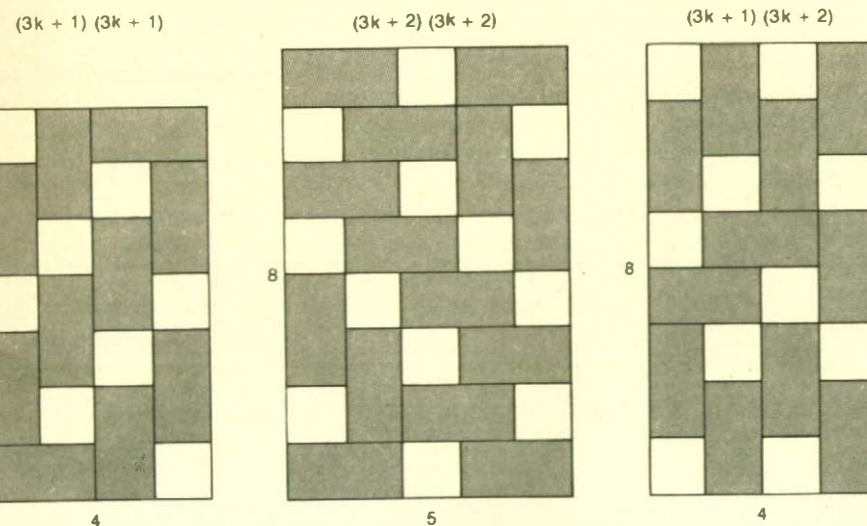
Le soluzioni della raccolta di brevi e disparatissimi problemi che sto per presentarvi saranno pubblicate il mese prossimo. Come al solito saremo felici di ricevere lettere che propongano correzioni, generalizzazioni, varianti, soluzioni migliorate o qualsiasi altro commento sui problemi. Non sarà possibile dare una risposta a tutte queste lettere, ma esse saranno vagliate attentamente e le questioni di particolare interesse in esse sollevate saranno riportate in uno dei prossimi articoli.

1. *Il problema di Gunport.* L'articolo del luglio scorso presentava un gioco detto cram in cui due avversari pongono a turno un pezzo del domino su una scacchiera rettangolare in modo che il pezzo ricopra esattamente due caselle, dopo aver convenuto se l'ultimo a muovere sia il vincitore oppure il perdente, come accade nel gioco alla rovescia. Len Gordon e Michael Beeler hanno richiamato la mia attenzione sul fatto che determinare la più lunga possibile delle partite di cram effettuabili su una scacchiera m per n equivale a risolvere un divertente problema di natura combinatoria tuttora irrisolto. Questo problema è stato presentato e fornito di un nome da Bill Sands nel suo articolo *The Gunport Problem* (« Mathematics Magazine », Volume 44, settembre 1971, pagine 193-196).

Il problema ha una formulazione semplice: qual è il massimo numero di « buchi » di lato unitario che può essere ottenuto disponendo i pezzi del domino su un campo m per n? Per ipotesi m ed n sono maggiori di uno.

Sands è riuscito a dimostrare che il numero dei buchi non può superare il numero dei pezzi del domino e che, se uno dei lati del campo è multiplo di 3, esiste uno schema che applicato ripetutamente permette di ottenere con facilità il massimo numero di buchi (*si veda la figura in alto in questa pagina*). In altre parole, se un lato del campo è pari a $3k$, il massimo numero di buchi è di $mn/3$, in caso contrario sarà un numero inferiore.

Murray Pearce di Bismark, North Dakota ha fatto l'ipotesi che il massimo numero di buchi sia uguale a $(mn-4)/3$ nel caso in cui i due lati, pur non essendo nessuno dei due uguale a $3k$, siano uguali modulo 3 (siano cioè entrambi o $3k+1$ o $3k+2$), e sia invece $(mn-2)/3$ quando un lato è della forma $3k+1$ e l'altro $3k+2$. Nella figura in basso di questa pagina sono riportati esempi di come possa ottenersi tale numero di buchi per i tre tipi di campi che non hanno un lato pari a $3k$.

Schema per gunport quando un lato del rettangolo è di 3k.

Schema per gunport quando nessun lato del rettangolo è di 3k.

Con le formule di Pearce il massimo numero di buchi per il campo 8 per 10 risulta essere 26. Sands ha però ammesso di non essere riuscito a ottenere più di 24 buchi usando 28 domino. Qualche lettore è in grado di trovare una soluzione con 26 buchi utilizzando 27 domino?

Naturalmente saranno i benvenuti tutti gli schemi che forniranno soluzioni con un numero di buchi superiore a quello previsto dalle formule di Pearce, come pure qualsiasi dimostrazione della correttezza delle due formule di Pearce.

2. *Le cifre non mentono.* Una vecchia farsa piuttosto nota presenta due sempliciotti che, dividendo 28 per 13, riescono a ottenere 7 e quindi eseguono la verifica del risultato ottenendo di nuovo 28 dalla moltiplicazione di 13 per 7 e infine, a coronamento della dimostrazione, addizionano 13 sette volte per ottenere 28. Ecco come Irvin S. Cobb ha raccontato la storiella nella sua antologia di 366 scherzi (uno per l'anno bisestile), *A Laugh a Day Keeps the Doctor Away* (Una risata al giorno leva il medico di torno).

« Tre aristocratici del deposito di carbone si erano messi in viaggio pieni di buone intenzioni, ognuno nel suo grande carro nero. Il loro sovrano, il sopraintendente al deposito, aveva dato ordine di distribuire in parti uguali a sette famiglie un totale di 28 tonnellate di carbone.

« Come furono usciti dal deposito, ognuno col suo primo carico, Kelly, Burke e Shea si fermarono a discutere il problema della ripartizione in parti uguali: quanto carbone toccava a ciascuna famiglia?

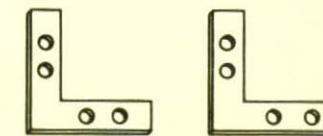
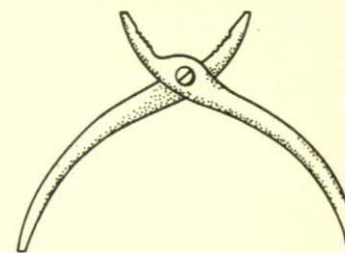
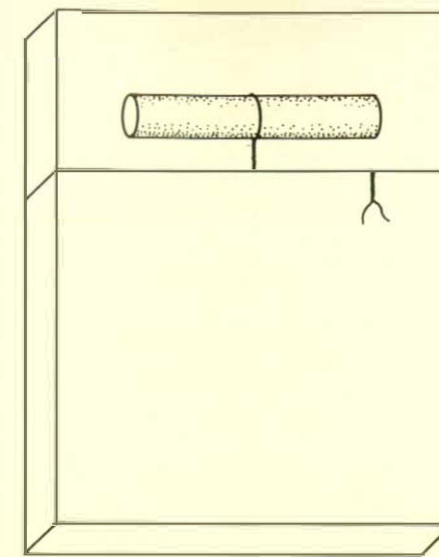
« "Bisogna fare così – disse Burke. – Non si tratta che di un piccolo conto. Se ci sono 7 famiglie e 28 tonnellate di carbone basta dividere 28 per 7, cosa che è presto fatta nel modo seguente: il 7 nell'8 ci sta una volta, e scrivo 1, il 7 nel 21 ci sta tre volte, e scrivo 3 ottenendo 13". Mostrò quindi trionfante il suo calcolo scritto con una matita smozzicata su un sudicio brandello di carta:

```
28 |7
 7  13
--
21
21
--
00
```

« Quelle cifre avevano fatto colpo, ma Shea non era del tutto convinto. "È facile vedere se è giusto, – dichiarò. – Basta sommare 13 per sette volte." Mise quindi in colonna le cifre come richiedeva la sua formula, e, partendo dalla base della colonna dei 3, arrivò in alto con un totale di 21 per ridiscendere lungo la colonna degli 1 contando così: "3,6,9,12,15,18,21,22,23, 24,25,26,27,28." Poté quindi annunciare in tono definitivo: "Burke ha ragione."

« Ecco il calcolo che Shea aveva mostrato:

```
13
13
13
13
13
13
13
--
28
```

« "Non sono perfettamente convinto – disse Kelly. – Lasciate che provi con il mio metodo. Se moltiplicando 13 per 7 ottenete 28 allora 13 è sicuramente il numero giusto." Tirò fuori dunque un mozzicone di matita e un foglio di carta dicendo: "Basta fare così: sette volte 3 fa 21, sette volte 1 fa 7 che sommato a 21 fa 28. È dimostrato così che 13 è la cifra giusta e quindi avete ragione entrambi. Volete vedere il calcolo?"

« Il parto del genio matematico di Kelly era il seguente:

```
13
 7
--
21
 7
--
28
```

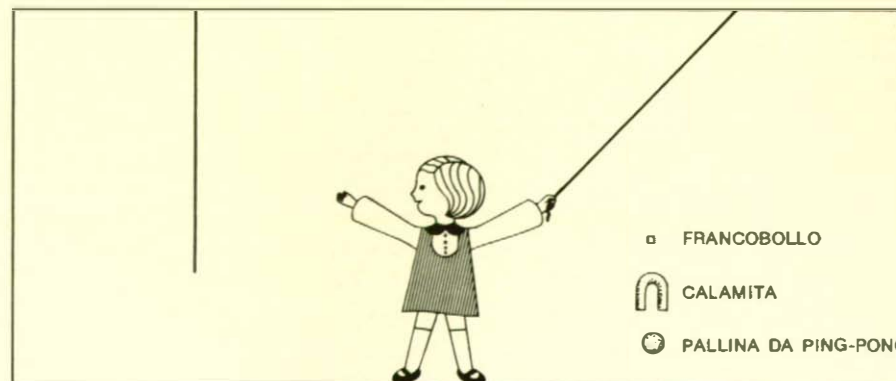
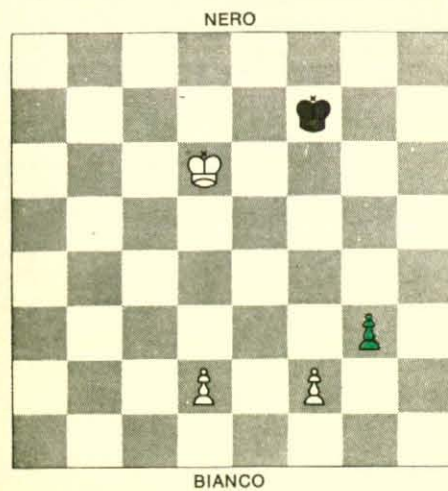
« "Non c'è altro da aggiungere, le cose stanno davvero così – dichiararono completamente d'accordo, e quindi distribuirono 13 tonnellate a ogni famiglia. »

Il commediografo Flournoy Miller ha effettivamente utilizzato questo sketch e nel suo libro *Shuffin' Along* ne ha pubblicato una versione. Pochi anni fa Flip Wilson fece uso di questa scenetta nel suo spettacolo televisivo e fu citato in giudizio dalla figlia di Miller per non aver rispettato le leggi sui diritti d'autore. Naturalmente il caso fu poi risolto in via amichevole.

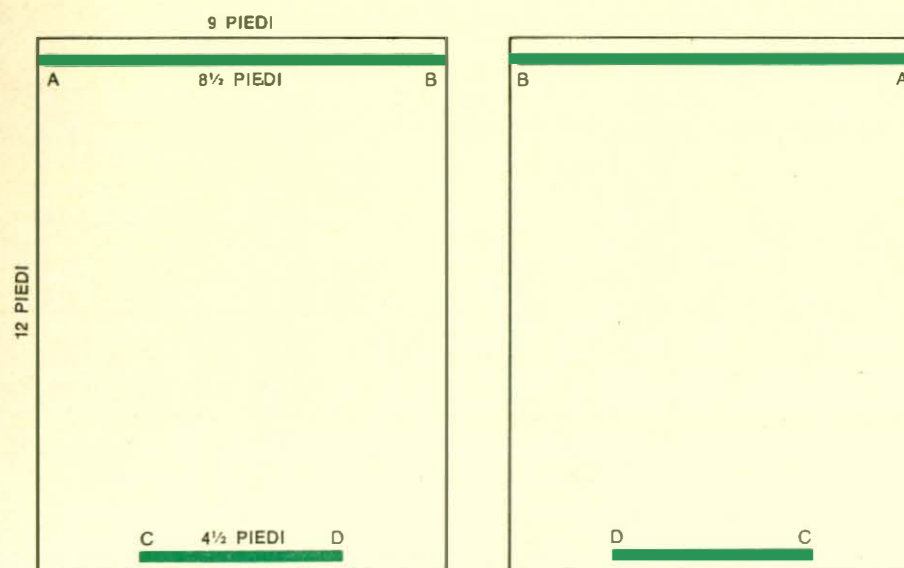
« C'è qualcosa di speciale nei numeri 7, 13 e 28? » si chiedeva il defunto William R. Ransom, un matematico della Tufts University. La risposta è no. Ci sono esattamente 22 triple di numeri che possono essere usate al posto di 7, 13 e 28 senza cambiare una parola del resto, in particolare in ogni tripla vi è un numero di una cifra e i restanti di due. Al lettore proponiamo di trovare queste 22 triple.

3. *Fissità funzionale.* L'esperienza vissuta può avere un effetto negativo sul pensiero creativo. Quando essa comporta una difficoltà nel vedere come un oggetto familiare possa essere utilizzato in maniera diversa dal solito, gli psicologi dicono di trovarsi di fronte a una manifestazione di fissità funzionale. Ecco due problemi, ben noti agli psicologi, che illustrano il concetto.

Siete seduti a un tavolo vuoto e vi vengono dati sei oggetti: un'assicella, un paio di pinze aperte al massimo, due piccoli angoli metallici con buchi per le viti, un piuolo e un tratto di

Il problema del piedistallo.

Il problema delle corde.

Il problema monocromatico di Smullyan.

filo di ferro che è stato utilizzato per legare strettamente il piuolo all'assicella (*si veda la figura in alto della pagina precedente*). Come potete disporre questi oggetti in maniera tale che l'assicella venga a formare un piedistallo orizzontale sopraelevato di qualche centimetro rispetto al piano del tavolo e abbastanza solido per reggere un vaso di fiori?

Siete in una stanza completamente spoglia dal soffitto della quale pendono due corde (*si veda la figura in basso della pagina precedente*). Il vostro compito è quello di legare insieme le due estremità libere delle corde e il problema sta nel fatto che, quando tenete un capo, l'altro, per quanti sforzi facciate, penzola sempre al di fuori della vostra portata. Non vi è permesso di usare alcun oggetto che abbiate con voi né alcun capo del vostro vestiario (come, per esempio, le calze, con cui potreste prolungare una corda per riuscire ad acchiappare l'altra), potete però far uso di uno dei tre oggetti posti sul pavimento: una pallina da ping-pong, una piccola calamita a ferro di cavallo e un francobollo.

4. *Scacchi monocromatici.* Molti lettori ci hanno comunicato il piacere col quale hanno accolto il brillante problema scacchistico, basato su una analisi « retrograda », inventato dal matematico Raymond Smullyan e presentato nell'articolo del maggio scorso. Smullyan, che sta ora preparando un libro di suoi originali problemi scacchistici inediti, ci ha permesso di presentarne un altro, appartenente a un gruppo di problemi da lui definito « scacchi monocromatici ».

La figura in alto ci mostra le posizioni di un finale di partita con soli cinque pezzi: il re bianco e quello nero, due pedoni bianchi e uno di colore ignoto (*in colore nella figura*). Sapendo che durante la partita non è stata compiuta alcuna mossa che portasse un pezzo da una casella di un colore a una di un altro, siete in grado di determinare se il pedone di colore ignoto sia bianco o nero?

5. *Le due librerie.* Robert Abes del Courant Institute of Mathematical Sciences dell'Università di New York è l'autore di questo problema che mi ha comunicato l'anno scorso. Una stanza di 9 piedi per 12 contiene due librerie in cui è collocata una preziosa collezione di romanzi erotici. La libreria *AB* è lunga 8 piedi e mezzo mentre la libreria *CD* è lunga 4 piedi mezzo. Le librerie sono entrambe collocate al centro della parete che occupano, a un pollice di distanza dalla parete.

I nipotini del proprietario devono venire a fargli visita e questi, facendo ruotare le librerie in modo che rivolgano verso il muro la loro parte anteriore, vuole evitare un brusco contatto fra bambini e libri. Ognuna delle due librerie deve trovarsi al termine della rotazione esattamente al punto di partenza, con le posizioni delle estremità invertite (*si veda la figura in basso*). Le librerie sono talmente pesanti che l'unica maniera di spostarle consiste nel sollevarle da una parte e farle ruotare secondo un arco di cerchio attorno all'altra estremità fungente da perno. Le librerie hanno una profondità molto ridotta e per gli scopi del problema è possibile rappresentarle in maniera idealizzata come dei segmenti di rette. Le estremità delle librerie, durante i loro spostamenti semicircolari, non possono attraversare le pareti né le librerie stesse. Qual è il minimo numero di rotazioni occorrente per voltare verso il muro le librerie?

6. *Probabilità irrazionali.* È molto facile usare una moneta come generatore di eventi casuali per decidere tra due possibili alternative dotate di probabilità espresse da numeri razionali. Supponiamo che vogliate scegliere tra *A* e *B* con una probabilità di 3/7 per *A* e una di 4/7 per *B*. Le possibili successioni di risultati ottenuti lanciando una moneta per *n* volte sono 2^n, per esempio tre lanci possono dare luogo a otto possibili successioni di risultati dati dalle triple: *TTT, TTC, TCT* e così via. Eliminiamo ora una di queste triple e scegliamone quindi tre a caso, stabilendo che siano esse a decidere per *A*, mentre le rimanenti quattro decideranno per *B*. Lanciamo la moneta tre volte: se il risultato è la tripla che abbiamo scartato ricominciamo da capo finché tre lanci consecutivi non diano una delle sette triple che abbiamo conservato. La probabilità che questa tripla appartenga all'insieme di tre triple è chiaramente di 3/7 e, analogamente, 4/7 sarà la probabilità che appartenga all'insieme di quattro triple.

La procedura è facilmente estendibile al caso in cui si debba decidere tra *n* alternative, ognuna con probabilità razionale. Supponiamo che vi siano tre alternative aventi probabilità $A=1/3, B=1/2, C=1/6$. Usiamo la tecnica precedente per decidere tra 1/3 e 2/3 (la somma di 1/2 e di 1/6). Se la decisione è per *A* vi arrestate, altrimenti proseguite decidendo tra *B* e *C*. A questo scopo dividete 1/2 (la probabilità di *B*) per 2/3 ottenendo 3/4 e dividete 1/6 (la probabilità di *C*) per 2/3 ottenendo 1/4. Come nel caso precedente si userà la moneta per decidere tra $B=3/4$ e $C=1/4$. La procedura si generalizza in maniera evidente al caso di *n* alternative, supponendo che si tratti sempre di probabilità razionali.

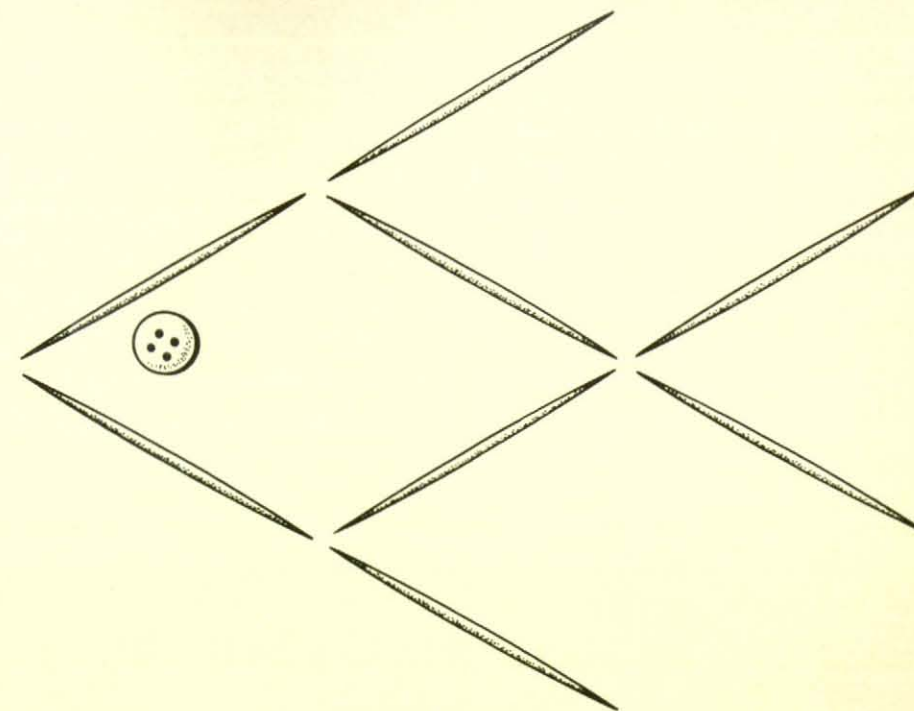
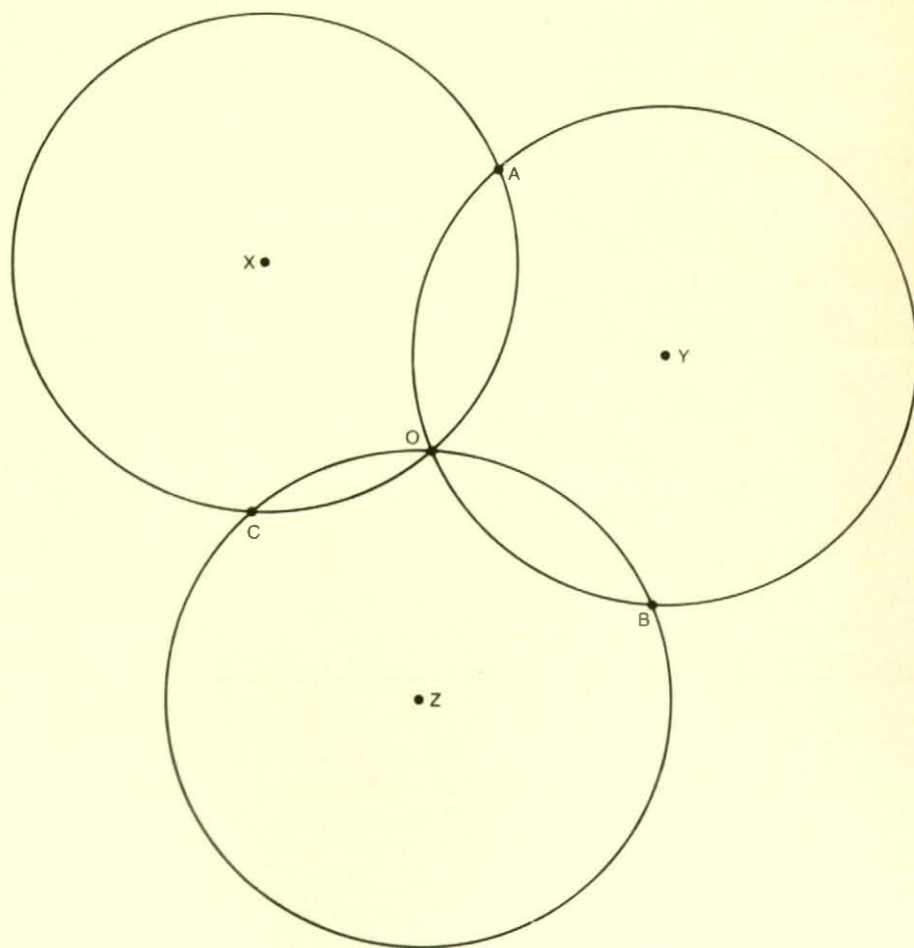
Inoltre non è necessario che la moneta sia perfetta. Supponete di avere a che fare con una moneta truccata e che venga testa con una probabilità di $1/\pi$: la probabilità che dopo testa venga croce resta uguale a quella che dopo croce venga testa. Non dovete far altro che tirare la moneta a gruppi di due volte e trascurare i risultati *TT* e *CC*, contando invece come croce i risultati *CT* e come testa quelli *TC*. Con queste due nuove definizioni di testa e croce, di eventi cioè equiprobabili, la moneta truccata può chiaramente usarsi tranquillamente per decidere tra *n* alternative, ognuna con probabilità razionale.

Supponiamo ora che vogliate decidere tra *n* alternative, ognuna con probabilità *irrazionale*. Sia per esempio *A* uguale alla parte decimale della radice quadrata di 2, *B* uguale alla parte decimale di π e *C* sia uguale a 1 meno la somma di *A* e *B*. Se potete decidere tra due probabilità irrazionali con una moneta perfetta potete farlo anche con una moneta truccata, ridefinendo testa e croce come abbiamo mostrato; potete inoltre decidere tra *n* alternative irrazionali quando siete in grado di decidere tra due, utilizzando il metodo fornito per *n* alternative razionali.

Ma il problema è appunto come una moneta possa essere utilizzata per decidere tra due probabilità irrazionali. Mettiamo a fuoco la questione con un esempio preciso. Sia $A=0{,}1415926535...$, cioè la parte decimale di π, e sia $B=0{,}8584073464...$, cioè 1 meno A: dovete decidere tra *A* e *B* lanciando una moneta. Persi Diaconis, un esperto di statistica dell'Università di Harvard, ha scoperto di recente un metodo delizioso per ottenere questo risultato, valido per tutte le frazioni irrazionali. Il mese prossimo ve lo riveleremo. (Suggerimento: il metodo fa uso della numerazione binaria.)

7. *Chi c'è dietro al Cappellaio Pazzo?* « Word Ways », rivista trimestrale di linguistica ricreativa, è arrivata al settimo anno di vita. I lettori che amano i rompicapo basati sulla lingua inglese possono abbonarsi inviando $ 7 al direttore, A. Ross Eckler, Spring Valley Road, Morristown, N.J. 07960. Il seguente problema carrolliano, proposto da John F. Collins di Santa Monica, California, è apparso nel numero di agosto del 1968.

« La Lepre di Marzo e il Cappellaio Pazzo stavano sorseggiando il loro eggnog, osservando la gente, quando Alice, dopo aver gettato uno sguardo nella direzione della Lepre, disse: "Per-

Il problema delle librerie.

Il rompicapo degli stuzzicadenti.

Il teorema delle circonferenze intersecantisi.

ché mi hai lanciato un'occhiata così irritata?"

« "Non te l'ho data, te l'ho semplicemente restituita" rispose la Lepre.

« "Ma io non ti ho guardato con irritazione."

« "D'accordo, ma qualcuno l'ha fatto" replicò la Lepre volgendosi verso il Cappellaio.

« Proprio in quel momento qualcuno sbucò da dietro il Cappellaio, chiudendogli gli occhi con le mani.

« "Indovina chi è!" Disse il nuovo arrivato con una voce sottile e incolore.

« Il Cappellaio si irrigidì per un attimo, poi dichiarò con una certa freddezza, "Non apprezzo gli scherzi pesanti."

« "Ah! Neanch'io" rispose lo sconosciuto, continuando a tenere le mani sugli occhi del Cappellaio.

« Sembrò allora che il Cappellaio accettasse la sfida, cominciò infatti a porre una serie di domande in un tono che lasciava trasparire speranza e apprensione.

« Domanda: "Ehm. Questa sera, per caso, sarà vestito di nero?"

« Risposta: "Non per caso ma per un preciso disegno."

« D. "Immagino che abbia una passione per i fiori."

« R. "Temo di no. Non li ho mai potuti soffrire."

« D. "Sarà certo al di sopra della media."

« R. "Ma è naturale!"

« D. "Non avrà delle macchie, spero."

« R. "Tocca ferro."

« D. "È sposato?"

« R. "No, felice." »

Chi c'è dietro al Cappellaio Pazzo?

8. *Rivoltate il pesce.* Questo affascinante rompicapo per bambini, che ho scoperto di recente in un libro di rompicapo giapponesi di Kobon Fujimura, è ben noto in Giappone ma pressoché sconosciuto da noi. Disponete otto stuzzicadenti e un bottone come è indicato dalla figura in alto della pagina precedente. Provate a invertire la direzione in cui nuota il pesce spostando solo tre stuzzicadenti e il bottone.

9. *Le circonferenze intersecantisi.* Si tratta di uno di quegli eleganti teoremi di geometria piana vecchio stile che a prima vista sembrano troppo difficili a dimostrarsi, ma a cui si arriva invece rapidamente con la giusta intuizione. Sono date tre circonferenze di raggio unitario e di centro X, Y e Z, che si intersecano tutte in un punto O (*si veda la figura in basso della pagina precedente*). Si dimostri che gli altri tre punti di intersezione A, B e C giacciono su una circonferenza, anch'essa di raggio unitario. Il problema è dovuto a Frank R. Bernhart, un matematico dell'Università di Waterloo in Canada.